Handling data insufficiency in semantic segmentation of hyperspectral images using deep neural networks: A review

Naftaly Wambugua, Zhenlong Xiaoa, Yiping Chena\*, Xiaoxue Liua, Cheng Wanga, Jonathan Lia, c \*

*a Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen, FJ 361005, China*

*c Department of Geography & Environmental Management and Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

**Abstract**

Over the years, advancements in sensor technology have improved spatial, temporal, spectral and radiometric resolutions, there by dramatically enhancing image scale, resolution and efficiency. These comprehensive advances have inspired improvements in various applications for hyperspectral image classification, such as land cover mapping, classification of vegetation, urban monitoring and understanding, all of which are important for better earth resource utilization. Processing these images, on the other hand, requires superior algorithms with higher precision, lower computational requirements, and robustness. Deep learning has revolutionized image-processing tasks and other new tasks. New deep convolution neural network (DCNN) models with superior image processing tasks capabilities are being proposed. However, in many studies, inadequate training data in hyperspectral images has been described as a significant bottleneck for most of the success of DCNNs in semantic segmentation. We provide a thorough analysis of the available datasets and the approaches that have handled training data insufficiency in DCNNs in order to simulate future studies. Our aim is three-fold: first, we present several available imagery datasets used in image processing domains with pointers to their core references; second, we examine how hyperspectral image (HSI) processing tasks have become improved through the generation of synthetic datasets, data augmentation, transfer learning, domain adaptation, and network optimization amid the glaring challenge on limited labeled data;. finally, we compare how different deep learning strategies were conducted on selected state-of-the-art image classification datasets and HSI assignments, and present the latest HIS semantic segmentation research directions.

***Keywords:*** Hyperspectral, Semantic segmentation, Remote Sensing, Data augmentation, Synthetic Data.

*∗Corresponding author at: Xiamen University*

*E-mail address: chenyiping@xmu.edu.cn (Yiping Chen) and junli@uwaterloo.ca (Jonathan Li)*

**1．Introduction**

Hyperspectral imaging (HSI) is a versatile technique that adds a new dimension to optical imaging. Researchers have been able to achieve high-quality spectral and spatial image resolutions with the availability of low-cost sensors, enhancements in image acquisition, processing techniques and sophisticated computing capabilities. In the recent past, it has grown increasingly popular to adopt hyperspectral imaging in various domains. The ability to precisely characterize the color of viewed item. What once required large, delicate and expensive laboratory spectrometers is now achieved using real-time aboard satellites, unmanned aerial vehicles and portable handheld units which stream large volumes of data in real-time with high precision and detail. Furthermore, processing this hyperspectral imaging big data in real-time demands superior computing power and algorithms(Fu et al., 2017a). One of the areas where hyperspectral imaging is widely applied is remote sensing for understanding the ground surface, it morphological changes and human processes with greater precision and detail.

**Semantic segmentation also referred to as pixel-level image classification (Liu et al., 2019a; Thoma, 2016) uses algorithms to assign each pixel in an image to a predefined class label where same-labeled pixels have similar characteristics (Kemker et al., 2018). Hyperspectral image (HSI) classification has several applications such as land cover mapping (Xu et al., 2019; Xu and Huang, 2014; Zhang, 2018), vegetation classiﬁcation (Laliberte et al., 2011), and urban monitoring and understanding (Huang et al., 2018a; Rottensteiner et al., 2012; Volpi and Ferrari, 2015; Zhang et al., 2015), among others. Pixel-based and object-based methods are some of the former approaches that focused on low-level features such as encoding spectral, textual, and geometric features to classify images and objects (Gaetano et al., 2015; Pesaresi and Gerhardinger, 2011; Rizvi and Mohan, 2011). These methods perform well for some specific problems but are notably poor in many other cases. They have been reported to demonstrate inferior capabilities in performing complex tasks such as image classification for HSI imagery which is characterized by noise, occlusion, shadows, illumination, season, and many other factors that make classification tasks using HSI images more challenging.** Castelluccio *et al.,* 2015 states that the image processing tasks get more complex as the level of abstraction shifts from the pixel, to objects and then to scenes. These image processing methods cannot handle most HSI segmentation tasks because hyperspectral images possess much higher spectral information than the RGB and other generic images. Each image pixel is a high-dimensional vector containing hundreds of spectral channels covering more than a single interval of wavelength range ( the wavelength range is typically in the range of 400-2500 nm) (Varshney and Arora, 2004).

The Hyperspectral data can be analyzed from either spatial, spectral, or joint spatial-spectral perspective. Most of the early Deep Learning(DL) methods only exploit data pixel-wise (1-dimensional approaches), working in the spectral direction. This is done by extracting spectral signatures from single pixels or groups of pixels either surrounding a central pixel or belonging to a given point of interest. This approach requires some prior knowledge and a pre-processing step to detect and map the regions of interest (usually done through segmentation). On the other hand, spectral-spatial classifiers integrate both spectral and rich spatial features to boost classification performance (Fauvel et al., 2008).

The artificial neural networks (ANN) concept was inspired by the design of the biological brain and tries to reproduce some of its functions using simple but massively interconnected units called neurons (Castelluccio et al., 2015). Recently, DL has proven superior compared to other shallow machine learning methods such as support vector machine (SVM), random forest (RF), and others that incorporate only low-level feature learning.

Deep learning (DL) is an ANN’s data-driven paradigm that provides an end-to-end machine-learning model that can extract salient features without human-designed algorithms. Feature extraction from lower layers is accomplished through simple non-linear model input that is channeled through some layers where the mapping relationship in these layers reduces the image dimensions and extracts key features of interest (LeCun et al., 2015). Intense research over the last few years has proposed a plethora of algorithms aimed at capturing more complex features and descriptors necessary to capture the semantics of the imagery for in-depth understanding using both spatial and spectral features (Chen et al., 2016b; Chen et al., 2014c; Chen et al., 2015b; Li et al., 2019a; Yue et al., 2015). To further enhance the classification performance of HSI, the development of more powerful models is necessary.

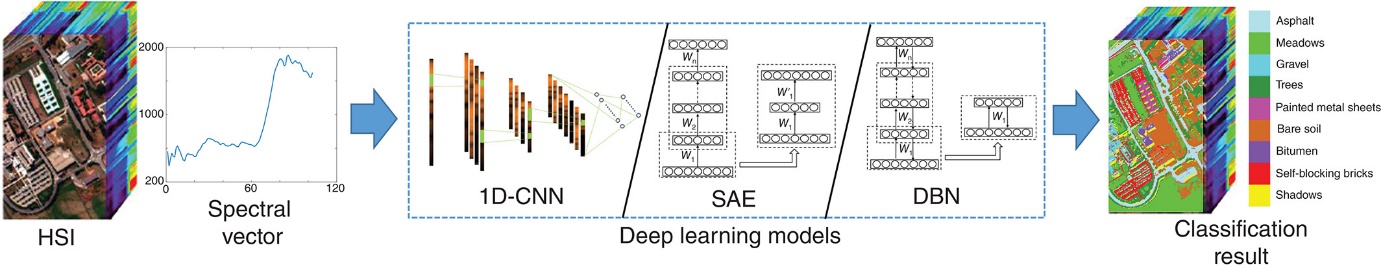
One of the key challenges faced in HSI semantic segmentation is the availability of limited label data. It is one of the key ingredient for the development of superior deep learning models. Mnih *et al.,* 2015 state that constructing a suitable and effective HSI dataset for model training needs for deep networks is expensive, time-consuming and offers a great challenge. Table: lists various datasets currently available for training deep learning networks, there is growing demand for Deep Learning Convolution Neural Network (DCNN) models for diverse domains and tasks.

To address the challenge of DCNN datasets, several methods have been proposed to handle labeled data insufficiency to improve deep learning models’ performance. One of them being to use transfer learning method, where the DCNN network is at first trained on large generic image dataset such as ImageNet. Several other large generic datasets outline in section 2.1 below also have a huge number of labeled training images and which are fine-tuned for other image-related tasks using the target dataset (Kemker et al., 2018).

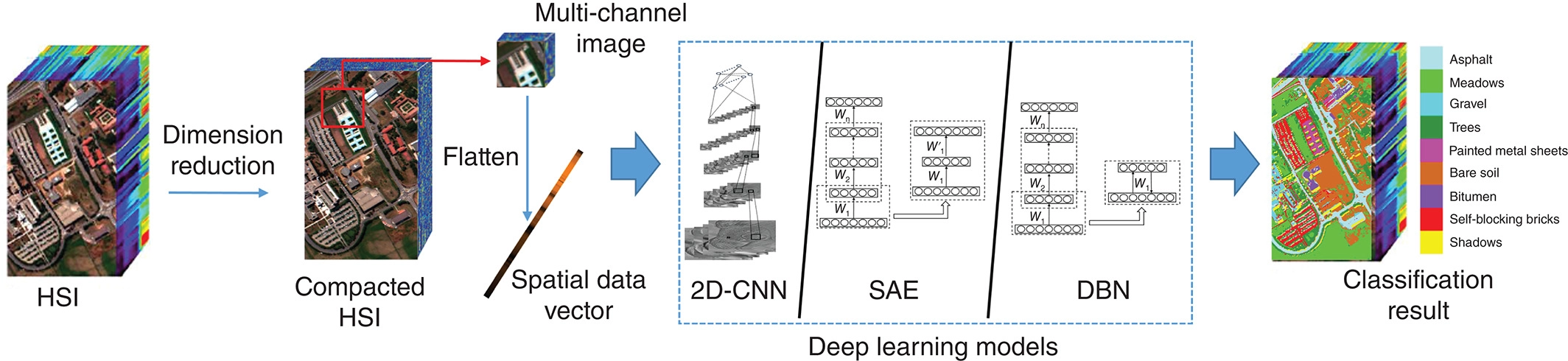
* 1. *Hyperspectral image semantic segmentation*

Recent HSI acquisition devices provide high spectral resolution images with sufficient spatial resolution for a variety of high-quality demanding applications (Khan et al., 2018). This presents a great opportunity for researchers to better understand the Earth’s changes and processes and the human activities on the Earth’s surface. However, the *curse of dimensionality* (Bellman, 2015), related to the great number of channels on the HSI images, and to the fact that data distribution becomes sparse and difficult to model as soon as the space dimensionality increases causing classification performance to deteriorate (Gao et al., 2014; Prasad et al., 2012). This happens especially when the number of available labeled training samples is limited thus limiting its exploitation and application (Li et al., 2013c). However, dimensionality reduction approaches have effectively been used as a pre-processing step allowing the removal of redundant features while preserving useful information in a low-dimensional subspace. Some of the dimensionality reduction techniques applied in HSI image classification tasks are:

* Principle Component Analysis (PCA)(Rodarmel et al., 2002), which is a classic method in countering the high-dimensionality cost,
* the maximum-noise-fraction (MNF) transform and supervised approaches like linear discriminate analysis (LDA), and
* local Fisher discriminate analysis (LFDA) (Li et al., 2011; Li et al., 2013b). PCA seeks to find a linear transformation through maximizing the variance in the projected subspace, whereas LDA tries to capture the inter-class and the intra-class variance.



(a)



(b)

Fig. 1. General framework of deep learning for spectral (a) and spatial (b) feature classification. (Zhang et al., 2016b)

Performing the dimensionality reduction while preserving the rich information content is a challenging task due to the complexities of high inter-class and low inter-class variations. Equally, the variable target boundaries and spatial distributions, and mixed pixels due to the images' fine detail, shadow, and occlusion complicate the image processing tasks such as segmentation and classification (Cheng et al., 2017). Fig. 1 illustrates the process of high-dimensionality reduction in HSI images before being fed into classical deep learning models for both spectral and spatial feature classification.

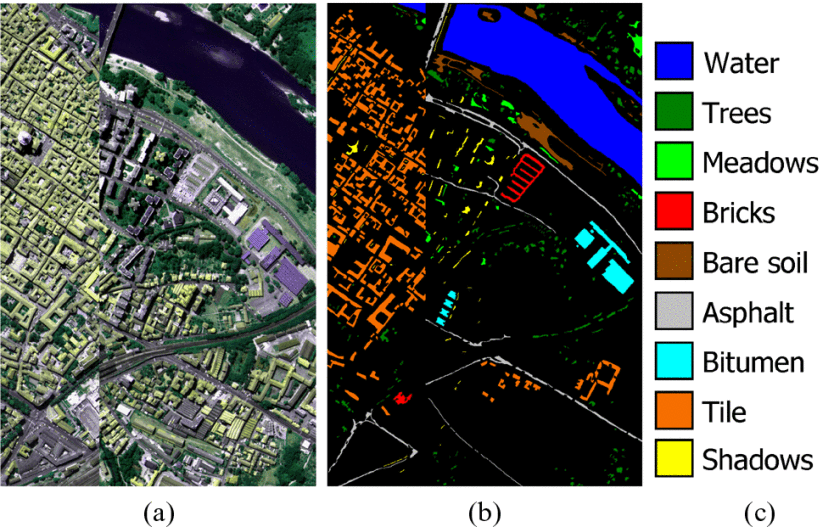


Fig. 2. A hyperspectral image of Pavia Center image with (a) Three-band false-color composite. (b) Reference data and (c) Color code. (Graña et al., 2013-2018).

Figure 2 illustrates a hyperspectral image of Pavia Center obtained from Reflective Optics System Imaging Spectrometer (ROSIS). The image spectral range is between 430 to 860 nm with 115 bands of 4.0 nm width and a spatial resolution of 1096 × 1096 pixels.

* 1. *Purpose of this review and relationship with other review works*

We present this review to highlight the approaches and methods proposed to handle labeled data insufficiency in semantic segmentation of hyperspectral images (HSI) using DCNNs. We discuss in detail the generic datasets used to pre-train models for later adaptation to HSI semantic segmentation tasks and thereafter describe the current trends in HSI segmentation. While this review is not meant to gain further insight into technical aspects of data acquisition mechanisms, data enhancement, and HSI applications, it exposes the research community to fundamental issues and trends relating to hyperspectral image processing and classification using DCNNs, and the methods used to handle insufficient labeled data. Highly informative reviews about DL methods in the RS ﬁeld have been produced focusing on theories, data analysis, and tools (Ball et al., 2017; Dua et al., 2020; Ma et al., 2019; Zhang et al., 2016a; Zhu et al., 2019; Zhu et al., 2017) where several references or sections focuses on HSI data. Different from other latest review work on deep learning on remote sensing application and HSI semantic segmentation, our work amasses various dataset categories (including generic datasets which are often used to pre-train HSI classifiers) with pointers to their core references to easily guide future researchers. We have also extended our work and highlighted promising approaches and recent methods that have effectively handled insufficient labeled data for image-related tasks and reported their performance. We have also highlighted the current methods that seek to optimally utilize the limited training data through robust network layers that can capture and extract robust descriptors using multi-layered and multi-leveled methods, Generative Adversarial Networks (GANs), and guided by attention mechanism among other approaches.

We have primarily focused on relevant recent papers from major conference repositories such as the Conference on Computer Vision and Pattern Recognition (CVPR), European Conference on Computer Vision (ECCV), Association for the Advancement of Artiﬁcial Intelligence (AAAI) Conference, International Conference on Computer Vision (ICCV), as well as Google Scholar. This review is intended to provide a detailed summary for researchers interested in evaluating various datasets useful for hyperspectral image classification, semantic segmentation and further building on methods and algorithmic techniques applied to them.

In this paper, we first highlight the hyperspectral imagery. Secondly, we present the available image datasets of different categories and demonstrate how the generation of synthetic datasets and data augmentation can be used to complement data insufficiency. We later discuss how transfer learning, semi-supervised feature learning, and deep domain adaptation seeks to learn features from the training data and re-use the features in related cases optimally with minimal data and how network optimization exploits parameters and network settings to make the deep network more robust and efficient in its functions. Finally, we report the performance of selected deep learning methods on semantic segmentation tasks in select public datasets and outline current research directions in the stated topic. (For clarity, in this report, we use “semantic segmentation” and “image classification” interchangeably to mean *pixel-level image classification*).

1. **Image datasets**

Data is one of the most critical components for any machine learning system and is increasingly critical for all computer vision tasks that use deep learning approach. Castelluccio *et al.,*(2015) state that the existing datasets are insufficient for training deep networks for HIS tasks and cause overfitting. While Mnih *et al.,* (2015) argues that the cost of constructing effective training datasets for HSI tasks are not only time consuming, but also require domain expertise and tend to be very time consuming. Alternatively, the use of standard image datasets may not generalize well due to the unique characteristics of HSI imagery (Kouw and Loog, 2019).

Existing datasets widely adopted for RS land cover classifications are Salinas, Pavia, Indian Pines and Kennedy Space Center. Signoroni *et al.,* (2019b) states that using these datasets can result in bias and not serve in improving the accuracy. In the subsequent sections, we discuss the use of generic image datasets which are widely adopted for hyperspectral Convolution Neural Network (CNN) classifiers as they transfer knowledge from well-labeled data sets to sparsely labelled datasets (Windrim *et al*., 2018).

* 1. *Generic image datasets*

A widely adopted strategy for training CNNs is the use of generic image datasets which are less time-consuming and easier to train a new classifier. Table 1 illustrates several datasets categorized by their application domains.

Table 1: Generic image data sets categorized by their application domain.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset Name | Domain | Classes | Training | Validation samples | Test | Year |
| ADE20K  (Zhou et al., 2017) | Generic | 150 | 20210 | 2000 | - | 2016 |
| COCO Stuff  (Caesar et al., 2018) | Generic | 172 |  | 163957 |  | 2017 |
| CIFAR - 10/100  (Doon et al., 2018) | Generic | 10 - 100 | 50K 500 | - | 10K/100 | 2009 |
| ImageNet  (Deng et al., 2009) | Generic | 1000 |  | 14,197,122 | 50K | 2010 |
| Apolloscape Scene Parsing Dataset  (Huang et al., 2018b) | Urban Understanding | 25 |  | 146997 Frames |  | 2018 |
| Barcelona  (Tighe and Lazebnik., 2010) | Urban Understanding | 170 | 14871 | - | 279 | 2010 |
| CamVid  (Brostow et al., 2009) | Urban understanding | 32 |  | 701 |  | 2009 |
| KITTI-LAYOUT (Alvarez et al., 2012) | Urban understanding | 3 | 323 |  |  | 2012 |
| KITTI-ROS  (Ros et al., 2015) | Urban understanding | 11 | 170 |  | 46 | 2015 |
| CityScapes  (Cordts et al., 2015) | Urban Understanding | 30 | 2975 | 500 | 1525 | 2016 |
| CBCL StreetScenes  (S, 2007) | Street view | 9 | 3547 |  |  | 2007 |
| Large-Scale Point Cloud Classiﬁcation Benchmark  (Hackel et al., 2018) | Urban/Nature | 8 | 15 |  | 15 | 2016 |
| INRIA-Graz-02 (Hackel et al., 2017) | Outdoor Natural | 3 | 479 |  | 479 | 2007 |
| DAVIS  (Pont-Tuset et al., 2017) | Videos | 4 | 4219 | 2023 | 2180 | 2016 |
| SegTrack v2  (Li et al., 2013a) | Generic video | 14 |  | 947 Frames |  | 2013 |
| Daimler Urban Segmentation (Scharwächter et al., 2013) | Street View video | 5 |  | 5000 |  | 2013 |
| MNIST  (Keysers, 2007) | Hand-written digits | 10 | 60000 | 28 by 28 pixel | 10000 | 2007 |
| KITTI-ZHANG  (Zhang et al., 2015) | Mobile robotics and autonomous driving | 10 | 140 |  | 112 | 2015 |
| GTSRB  (Stallkamp et al., 2011) | Traffic lights | 43 | 50,000 |  | 2 | 2011 |
| Mapillary Vistas (Neuhold et al., 2017) | Street view | 66 | 18000 | 2000 | 5000 | 2017 |
| Microsoft COCO  (Lin et al., 2015) | Image recognition, segmentation and captioning | 80 | 82783 | 40504 | 81434 | 2014 |
| LabelMe  (Russell et al., 2008) | Outdoor | 8 | 2920 |  | 1133 | 2008 |
| Sift-Flow (Liu et al., 2009) | Outdoor | 33 | 2688 |  | 200 | 2011 |
| Stanford Background (Gould et al., 2009) | Outdoor | 8 | 725 |  |  | 2009 |
| Stanford 2D-3D (Armeni et al., 2017) | Indoor / 2D-3D | 13 |  | 70469 Scans |  | 2017 |
| ScanNetv2  (Dai et al., 2017) | Indoor 3D | 20 |  | >1500 Scans |  | 2018 |
| PASCAL VOC  (Chen et al., 2014a) | Competition Challenge | 21 | 1464 | 1449 |  | 2012 |
| PASCAL PART  (Chen et al., 2014a) | Competition Challenge | 21 | 1464 | 1449 | 9637 | 2011 |
| PASCAL CONTEXT (Mottaghi et al., 2014) | Competition Challenge | 540 | 10103 |  | 9637 | 2014 |
| Cornell RGB-D  (Liu and Fan, 2013) | Generic applications |  | 8498 | 2857 |  | 2011 |
| RGB-D Object v2  (Lai et al., 2011) | Household objects | 51 | 207920 | 41877 |  | 2011 |
| Adobe’s Portrait Segmentation  (Shen et al., 2016) | Portraits | 2 | 1500 | 300 |  | 2016 |
| (URSA) Dataset (Angus et al., 2018) | Autonomous driving road parsing. |  |  | 1,355,568 road scenes |  | 2018 |
| YouTube Dataset  (Jain and Grauman, 2014) | Objects / Video | 10 | 10167 |  |  | 2014 |

* 1. *Remote sensing and aerial datasets.*

One of the key challenges faced today is the lack of adequate training datasets that would serve as representative quality benchmark for implementing new techniques in remote sensing domain. In addition to this, we are currently facing with capturing, analyzing large volumes of real-time data that not only demands higher computing requirements but also requires us to efficiently understand and predict land-cover and morphological changes using machine learning techniques.

*Table* 2 outlines various publicly available remote sensing and aerial datasets classified by their application domain. However, some of the datasets do not explicitly detail the image specification such as image size and spatial resolution.

**Table 2**: Remote sensing and aerial datasets classified by their application domain

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset Name | Application domain | classes | Training | Size / Sensor | Spatial resolution (m) | Year |
| SIRI-WHU Dataset  (Zhao et al., 2016a) | RS | 12 | 200 per class | 200 X 200 | 2 | 2016 |
| UC Merced Land-Use Dataset  (Yang and Newsam, 2010) | RS | 21 | 2,100 | 256 x 256 | 0.3 | 2010 |
| Brazilian Coffee Scene Dataset (Penatti et al., 2015) | RS | 2 | 1,438 | 64×64 |  | 2015 |
| Salinas Scene (Graña et al., 2013-2018) | RS | 16 | 111,104 | 83 x 83 | 3.7 | 1998 |
| RSSCN7 Dataset (Zhao et al., 2016b) | RS | 7 | 2,800 | 400 x 400 |  | 2015 |
| RSC11 Dataset (Zhao et al., 2016b) | RS | 11 | 1,232 | 512 x 512 | 0.2 | 2016 |
| Indian Pine  (Graña et al., 2013-2018) | RS | 16 | 10,366 | 145 x 145 | 20 | 1992 |
| WHU-RS19 Dataset  (Sheng et al., 2012) | Land cover | 19 | 1,005 | 600 x 600 | 0.5 | 2012 |
| AID Dataset  (Xia et al., 2017) | Land cover | 30 | 1,000 | 600x600 |  | 2017 |
| Botswana  (Graña et al., 2013-2018) | RS | 14 |  | 1496 × 256 | 30 | 2004 |
| 19‐Class (Xia et al., 2010) | RS | 19 |  | 600 X 600 |  | 2010 |
| SAT-4 Airborne Dataset  (Basu et al., 2015) | Land cover | 4 | 500,000 | 28x28 |  | 2015 |
| IEEE GRSS 2016  Data Fusion Contest  (Tuia et al., 2017) | RS |  |  | 3840×2160 | 4 | 2016 |
| SAT-6 Airborne (Basu et al., 2015) | Land cover | 6 | 405,000 |  |  | 2015 |
| RIT-18 | Semantic Segmentation | 18 |  |  | 0.047 | 2017 |
| SpaceNet (Vakalopoulou et al., 2017) | Building detection |  | 17,533 | 200 x 200 |  | 2017 |
| Forest Type Mapping Dataset  (Brian et al., 2012) | Trees classification | 4 | 326 |  |  | 2015 |
| Wilt Dataset  (Alan et al., 2013) | Understanding the forested environment | 2 | 4,339 |  |  | 2016 |
| TorontoCity  (Wang et al., 2016) | City understanding |  | 56,000 | 500×500m2 | 0.10cm/pixel | 2009 - 2013 |
| Freiburg Forest (Valada et al., 2017) | RS | 6 | 230 | - | 136 | 2016 |
| University of Pavia (Graña et al., 2013-2018) | RS | 9 | 42,761 | 610 × 340 pixels | 1.3 / pixel | 2001 |
| DLR’s Aerial Crowd Dataset  (Bahmanyar et al., 2019) | Crowd counting and density estimation. |  | 226,291 |  | 4.5 – 15 cm | 2019 |
| NWPU-RESISC45 (Cheng et al., 2017) | Scene Classification | 45 | 31,500 | 256×256 | ~30 to 0.2 | 2016 |
| xView Dataset  (Lam et al., 2018) | Object detection | 60 | 1,000,000 | 1,400km2 |  | 2018 |
| Deep Globe  (Demir et al., 2018) | Road extraction | 2 | 6,226 | 19,584×19,584 pixels | 50 cm/pixel | 2018 |
| Building detection | 8 | 10,146 | 20,448× 20,448 pixels | 50 cm/pixel |
| Land cover classification | 7 | 1,014 | 20,448× 20,448 pixels | 50 cm/pixel |
| Mnih dataset  (Mnih, 2013) | Toronto Roads |  | 500km2 |  | 1.2m/pixel | 2013 |
| Hamilton Roads Dataset |  | 250km2 |  | 1.2m/pixel |
| Zeebruges data  (Vo et al., 2016) | Land cover classification |  | 7 | 10,000 X 10,000 | 5 cm/pixel | 2016 |
| ISPRS2D Vaihingen semantic labeling dataset  (Gerke et al., 2014) | Land cover classification | 6 | 33 | 2,000 × 2,000 pixels |  | 2014 |
| ISPRS2D Potsdam semantic labeling dataset  (Gerke et al., 2014) | Land cover classification | 6 | 38 | 6,000 × 6,000 pixels |  |
| Inria Aerial Image (Maggiori et al., 2017) | Building detection |  | 810km2 |  | 30cm/pixel | 2017 |

* 1. *Synthetic datasets*

Minh *et al.* 2015 argues that the process of acquiring high-quality training samples to train deep learning networks for semantic labeling is time consuming. Hence, Patki *et al.,* (2019) proposes the use of simulation to generate a well-annotated dataset that is not only cheap, but also fast to create and is not constrained by the availability of time or the physics of the natural world. It is also important to bear in mind that there is no significant statistical evidence between the accuracy score of data scientists with control data and synthesized data (Patki et al., 2019; Rajpura et al., 2017; Tremblay et al., 2018). Over time, several DNNs have used synthetic datasets to perform computer vision tasks such as object detection (Pepik et al., 2012) and scene understanding (Satkin et al., 2012), while other researchers in the computer vision domain have used simulators to model objects such as human shape (Grauman et al., 2003), face and hand geometry (Ballan et al., 2012) and vision as inverse graphics (Battaglia et al., 2013) and yet obtained promising results. Compelling results were reported in (Hu et al., 2018), where real face images were composited to generate very large training datasets of synthetic images.

Among others, synthetic datasets have contributed to the vast advancement of computer vision and machine learning techniques such as segmentation, object identification, object recognition, object processing and semantic segmentation. Tobin et al., (2017) argues that using domain randomization the model interprets synthetic data as part of the training datasets which is indistinguishable from physical information, thus exposing the model to a wide range of environments during training (Peng et al., 2015). Furthermore, Synthetic datasets have helped to improve the model's performance and in some studies outperformed photorealistic datasets (Patki et al., 2019).

With the widespread use of synthetic datasets, Bousmalis *et al.,*(2018) argues that there is significant quality difference caused due to natural richness, non-rigidity and noise between original data and synthetic data and requires post-process validation. However, some users do not consider this as a valid reference model as it does not conform to any quality benchmarks. Rajpura *et al.,* (2017) argues that synthetic HSI datasets are suited where they exhibit high intra-class variance, high clutter, occlusion and where domain gap remains a key concern. Tobin *et al.,* (2017) has proposed domain randomization as one of the approaches to bridge the reality gap.

Table 3 outlines various synthetic datasets classified by various application domains.

**Table 3**: Synthetic datasets classified by the application domain.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset Name | Application domain | Classes | Size | Year |
| SYNTHIA  (Ros et al., 2016) | Urban scene understanding | 13 | 13,400 frames | 2016 |
| Data from Game  (Richter et al., 2016) | Street View | 19 | 24966 | 2016 |
| SceneNet RGB-D  (Mayer et al., 2015) | Indoor scene understanding | 13 | 5000000 | 2016 |
| NYUv2  (Silberman et al., 2012) | Indoor scene understanding | 13 | 1449 | 2016 |
| Synthetic Image Dataset (Bhoyar, 2014) | HSI Segmentation |  | 100 | 2014 |
| SUN3D (Xiao et al., 2013) | Indoor scene understanding | Video | 2.5 Million |  |
| Flying Chairs  (Dosovitskiy et al., 2015) | Indoor scene understanding | 809 | 22,872 | 2015 |
| Virtual KITTI  (Gaidon et al., 2016) | Video understanding | 5 Virtual worlds | 21,260 frames | 2016 |
| Driving Stereo  (Yang et al., 2019) | Estimating maps in autonomous driving |  | 180k |  |
| MPI Sintel (J et al., 2012) | Optical flow understanding | 3D Film | 23 training; 12 testing | 2012 |
|  |  |  |  |  |

1. **Handling data insufficiency**

A significant challenge in classification capabilities of most classifiers and algorithms is attributed to the availability of limited labeled datasets. In the subsequent sections, we discuss some methodology to improvise the image processing tasks with limited labeled datasets.

* 1. *Data augmentation*

A new technique to amplify and augment existing sample data is by applying transformations, sometimes referred as Data Augmentation. Similarly, image augmentation artificially creates training images through a combination of multiple processing approaches using color augmentations and geometric transformation such as shear, random rotations, shifts, flips on copies of existing images to improve robustness of the model (Wong et al., 2016). Tajbaksh *et al.,* (2019) argues that by applying traditional data augmentation methods the resulting output of the images are highly correlated with original images thereby limiting their impact.

As has been previously reported in the literature, two common methods are widely used to generate additional data samples:

1. Transformation-based sample generation
2. Mixture-based sample generation.

*Transformation-based sample generation:* Various authors (Aptoula et al., 2016; Chen et al., 2016a; Lee and Kwon, 2017; Xu et al., 2017) have proposed the creation of new virtual data using known samples by modifying their inherit characteristics. These HIS imagery suffer from complex and varied lightning conditions, similar class objects tend to appear at varying distance and perspectives due to different radiations and illuminations.

***Mixture-based sample generation*:** Same type objects in a given spectrum appear to have analogous spectral characteristics. Driven by this theory, it is possible to create virtual samples from two given samples of the same class by combining the samples of the same class linearly. Chen *et al.,*(2016a) and Kang *et al.,*(2018) built simulated samples using a linear combination of various training samples. These two approaches have proven efficient in enhancing the performance of the model in different tasks. Nalepa *et al..*(2019) investigated the power of data augmentation for deep networks on hyperspectral data at inference time and demonstrated it to be a promising generalization technique for deep learning networks.Several other authors (Li et al., 2018; Nalepa et al., 2020b; Wang et al., 2019) have proposed in their works various data augmentation methods such as : , which seek to increase the number of training samples in RS image classification tasks. Their works demonstrates that data augmentation is a cost-effective and simple way of handling labeled data insufficiency for image processing tasks, especially for HSI image segmentation.

**Very Recently, Buslaev *et al.,*(2020) presented albumentations library which is a fast and flexible tool for image augmentations with many various image transform operations relating to color, contrast, brightness, and other geometric transformations in high resolution \_\_\_\_\_(VHR)** (Liu et al., 2020a).

* 1. *Generative adversarial networks*

Generative adversarial networks (GANs) consist of two parallel parts that are both parameterized as deep neural networks. As shown in Fig. 3, a generator G produces synthetic data given a noise variable input Z while a discriminator D identifies whether a sample is coming from the real data distribution Xr or the generated data distribution Xg. The discriminator D is trained to estimate the probability of a given sample coming from the real data distribution whereas the generator G is optimized to “fool” the discriminator to oﬀer a high probability for the generated data (Luo et al., 2019b).

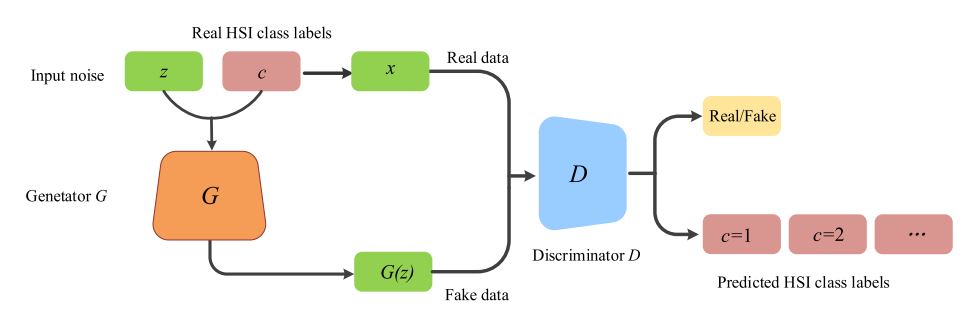


Fig. 3. An illustration of the GAN-based hyperspectral imagery (HSI) classification approach (Lin et al., 2018c)

In order to produce new image samples for training using the min-max strategy, GANs are considered an efficient unsupervised technique where one neural net successively produces false samples from the original data. GANs can learn how to create data that is indistinguishable from the original data from a dataset. A style image from a subset of choice of styles is chosen from each source image and a tailor-made transformation of the original image is generated. CycleGAN (Zhen-long et al., 2019) has been used for style transfer such as transferring images in one setting to another setting. GANs, for example, can be used to train a car in a night scene or a rainy scene using only data collected on these scenarios. These networks have been reported to work well even with limited datasets and have shown extremely good functionality in increasing image resolution of input images (Marchesi, May 2017). GANs have been used in various prior RS studies. Yun et al. (2019) trained cycle consistent adversarial networks (CycleGAN) to convert visible images to generate synthetic InfraRed images for training their network. Similarly, Benjdira *et al.,* (2019) used the output of CycleGANs between the visible band and infrared data to significantly increase the segmentation accuracy of RS datasets, while Seo et al. (2018) transferred image statistics from real images into synthetically rendered imagery containing military vehicles to increase the overall image conformity. In each of these works, real data were used to augment synthetic data for object detection or segmentation model training.

* 1. *Transfer learning*

One of the key characteristic of the deep learning model is to extract high-level features using training data and then further tune fine features to produce accurate prediction by using an optimization algorithm (Lundervold and Lundervold, 2019). Transfer learning techniques are widely adopted to read from the source dataset features and apply to the target dataset (Yosinski et al., 2014). This approach significantly decreases the demand for training samples, improves the models learning, and decreases the training time required. Usually, features learned from the lower layers of a CNN such as color blobs, edges, and other low-level features are well generalized for other classification tasks (Castelluccio et al., 2015). The main aim of transfer learning is to train deep networks using significantly less data than it would require if the network is to be trained from scratch. This allows machine algorithms to utilize the knowledge acquired from one task to solve related tasks.

Generic descriptors obtained by CNNs are powerful in that they can provide a generic image representation, the ability to tackle diverse tasks of object recognition, image classification, scene recognition, fine-grained recognition, attribute detection, and image retrieval applied to a diverse set of image datasets (Razavian et al., 2014). This makes transfer learning an astounding baseline for image processing tasks.

As shown in Fig. 4, the learnt features from low and middle network layers are transferred to other networks within the same architecture as the learnt one in an image segmentation task. After the transfer, the new network can use the learnt features for classification tasks using either supervised or unsupervised methods. During pre-training, the network learns salient features from the network to perform classification while fine-tuning the network is tweaked with a small number of training samples for the target task (Long et al., 2015b). During fine-tuning, some initial network layers are frozen because low-level features can best fit for most deep learning problems, while a few top layers are adapted to learn features of the target task. Transfer learning has been reported effective in improving network performance especially in cases where training data is insufficient (Yosinski et al., 2014).

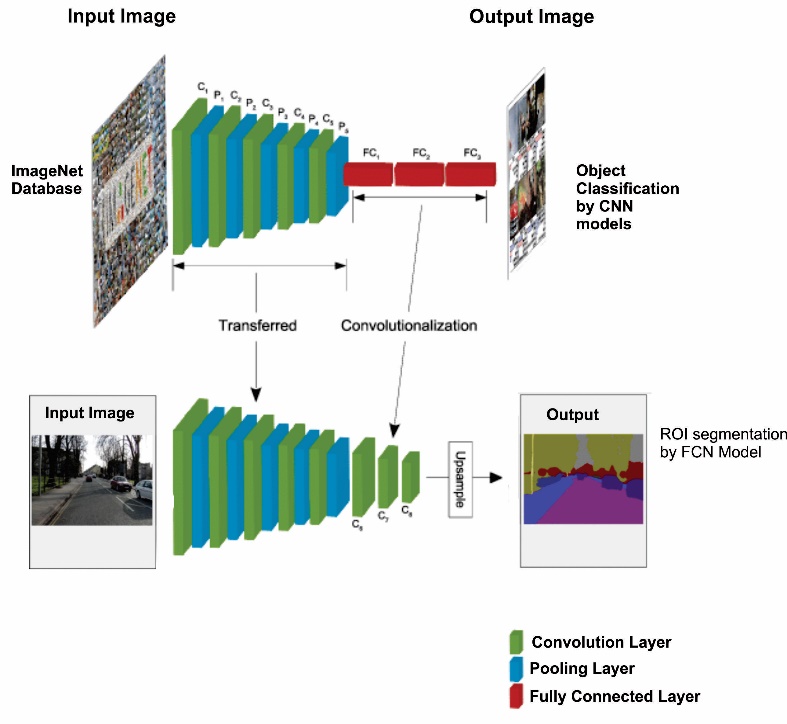
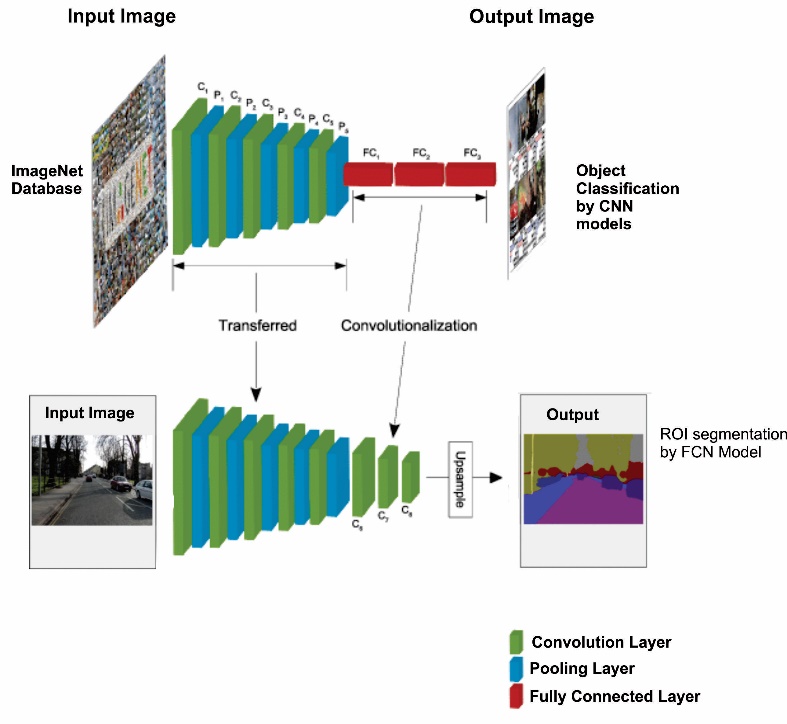


Fig. 4. Transfer learning procedure of deep CNNs to obtain optimized weights initializations (Ahmad et al., 2018).

* 1. *Domain adaptation*

In reality, sometimes researchers have adopted auxiliary data sets to compensate for quality training data for a given task. These auxiliary data sets have varying pose, illumination and image quality, sometimes there exists some domain variance between two application domains, which can lead to degrading performance (Wang and Deng, 2018).

Domain adaptation (DA) is a transfer learning paradigm that exploits labeled information in one relevant application to execute new tasks in another application domain. The goal of domain adaptation is to minimize the domain gap (where data from the source and target domains may either be considered homogeneous or heterogeneous) and to transfer knowledge in cases where a wide domain gap exists while attempting to reduce dataset bias caused by the difference in the statistical distributions between training and test domains (Hoffman et al., 2014).

In various cases, images collected for computer vision tasks may suffer platform inconsistencies due to diverse sensor settings, image acquisition platforms, image calibration settings, and device’s mechanical configuration variations. Consequently, images trained on one platform may fail to perform well in another platform thus requiring an adaptive system that seeks to bridge the difference between two environments (Tuia et al., 2016a). A classic example is training a model on SAR imagery versus an optical dataset. The two datasets fall on different domains and thus a model that has been trained on one model may not generalize well in another target task (Tuia et al., 2016b). UDA requires no prior knowledge on the label sets from a given source label set and a target label set. Sometimes, both the source and target domains may contain a common label set and hold a private label set respectively thus bringing up an additional category gap (You et al., 2019).

UDA leverage deep networks to learn more transferable representations by embedding domain adaptation in the pipeline of deep learning. It is different from the standard DA methods which attempt to learn the domain invariant feature representation by generating features for target domains or by transforming samples between domains through generative models. Unlike UDA, DA methods have been reported to heavily rely on data labels to establish a relationship between source and target datasets (You et al., 2019). DA in RS imagery on Hyperion, National Center for Airborne Laser Mapping (NCALM), and WorldView-2 datasets demonstrates the power of DA for HSI-RS imagery in (Song and Ma, 2017) and HSI image classification (Fan et al., 2019) using the unsupervised approach and how it effectively addressed the issue of limited or unlabeled datasets.

Researchers from different application domains using deep learning have proposed various approaches for domain adaptation techniques especially on how vision-based models trained in a given domain can fit in given target domain (Hoffman et al., 2014; Hoffman et al., 2013; Kulis et al., 2011; Li et al., 2014; Long et al., 2015b); how models can be re-trained in the target environment (Yosinski et al., 2014); how weights of a pre-trained can be adapted (Li et al., 2016); how to use pre-trained weights to extract features (Gupta et al., 2016); and how models can learn similar features between domains (Tzeng et al., 2014). Nogueira et al. exploited different strategies for exploiting the power of existing DCNNs in different scenarios from the ones they were trained on. They explored full training, fine-tuning, and using ConvNets as feature extractors and found out that fine-tuning outperforms the other approaches including in RS applications (Nogueira et al., 2016).

* 1. *Unsupervised / semi-Supervised feature learning*

Emerging CNN’s models aiming at addressing the complexities of imagery are being proposed constantly. These new networks are getting deeper and wider thus requiring massive labeled data to train them on. The unavailability of labeled datasets in HSI and related domains where the task of labeling imagery is more laborious and complicated opens the shift for semi-supervised feature learning.

Unsupervised/Semi-supervised feature learning methods either extract salient features from unlabeled datasets or uses unlabeled data during pre-training or labeled data for fine-tuning. Various research works have focused on designing deep learning frameworks for HSI-RS image classification (Mou et al., 2018; Radford and Metz, 2016; Romero et al., 2016; Tao et al., 2015). (Chen et al., 2014b; Chen et al., 2015a; Liu et al., 2017; Ma et al., 2016) adopted a fully connected network architecture using a semi-supervised approach to classify HSI imagery. Wu and Prasad (2018), employed a non-parametric Bayesian clustering algorithm to generate pseudo labels that help in pre-training Convolution Recurrent Neural Network (CRNN) for the HSI classification task, (Mou et al., 2018) proposed an end-to-end unsupervised feature learning using convolution and de-convolution networks in place of encoder and decoder, while (Zhan et al., 2018) employed GAN to design a semi-supervised feature learning framework for HSI classiﬁcation where the generator created counterfeit hyperspectral sample images that were similar to the real data to train the GAN. The current state of the art of unsupervised/semi-supervised feature learning assumes the encoder-decoder concept without using labels data and has shown capabilities to extract sufficient spectral and spatial features suitable for satisfactory results.

Over time, other research works have proposed other network optimization approaches to improve network generalization capabilities. (Salimans and Kingma, 2016b) regularized the pre-training and fine-tuning procedure and obtained superior classification results for Deep Belief Network (DBN) models, (Ba et al., 2016) enforced label consistency constraints into the training phase of stacked auto-encoders (SAE) while (He et al., 2016; Zhong et al., 2018) explored the correlation between training sample and their impact on network performance.

The above methods, (sections 3.1 to 3.5) were proposed to adequately handle the data-hungry DL models. Combining multiple and diverse architectures have been noted to improve the model’s generalization and robustness on diverse categories of images by extracting different levels of semantic representations thus achieving superior results (Sohail et al., 2020). Most of the methods discussed in section 3 have achieved compelling results in various image processing tasks with limited data labels.

1. **Network tuning and optimization methods**

Sifting through the current literature, it can be summarized that deep learning is promising research area especially in Deep Learning Convolution Neural Network (DCNN) design and configuration. They apply generalized optimization methods to extend the networks’ capabilities by tuning efficient functions and settings. They have been applied in various DCNNs including those designed for image classification tasks and have demonstrated superior results.

1. *Regularization techniques* are modifications to the learning algorithm that attempts to reduce the generalization error and not its training error by pushing the coefficients for many variables to zero and thus reducing the cost term (Ma et al., 2016). The net effect of regularization techniques can reduce overfitting challenges.
2. *Data augmentation* as discussed in section above, this method attempts to extend the dataset by creating geometric transformation on existing labeled data.
3. *Normalization* is a data pre-processing exercise where all data is transformed to be on the same scale to solve a situation where some numerical data points may be very high while others very low which causes a neural network difficult to train due to an exploding gradient problem. Classic normalization scheme scales numerical data down to a scale between 0 and 1 in a process that involves subtracting the mean of the dataset from each data point and then dividing that difference by the data set’s standard deviation.
4. *The batch normalization technique*, provides any layer in a Neural Network with inputs that are zero mean/unit variance. This allows better gradient flow during the network training phase by normalizing the layers and allowing the network to train the normalization weights. Weight normalization (Salimans and Kingma, 2016a) and layer normalization (Luo et al., 2018) are advanced normalization variants that seek to solve the dependency on the mini-batch problem of the batch normalization method (Ioffe and Szegedy, 2015).
5. *Residual learning technique* (Krizhevsky et al., 2012) addresses degradation problems (where accuracy gets saturated and then degrades rapidly with an increase in network depth). By adding identity mappings to CNN, deep residual networks (ResNets) mitigates the declining-accuracy effect and offers better results.
6. *The dropout technique* is inspired by the fact that the more capacity you add to the model, such as more layers and more neurons, the more the network becomes prone to over-fitting. Drop out works by using probability to remove a neuron from a given network layer during training or by dropping certain connections that have a high probability of causing dead-ends (Srivastava et al., 2014). Dropout is used together with other techniques like L2 regularization.
7. **Deep learning methods for semantic segmentation**

It has taken many years to attain remarkable levels of performance in the current domain of semantic segmentation. CNNs performance on image processing tasks marks notable progress that has increasingly advanced from the early architectures. To handle the challenges associated with HSI images, deep architectures designed for the category and pixel-wise labeling must be able to delineate objects and boundaries accurately between various classes and maintain high spatial resolution during classification. Equally, due to the nature of remotely sensed data, models designated for RS domains should be able to effectively discriminatively encode spectral and spatial information (Nogueira et al., 2016).

Most of the DCNNs are based on hierarchical feature learning and are trained in an end-to-end framework in which high-level features and low-level features are learnt hierarchically (Alvarez et al., 2012; Brostow et al., 2009; Huang et al., 2018b; Tighe and Lazebnik., 2010). Feature maps at different levels of abstraction present a complex mapping function from input raw data and outputting the desired class. Such complex mapping not only considers the spectral information of every pixel but also puts into account the textual, contextual, and spatial information.

In this section, we explore how deep learning methods learn feature representation for semantic segmentation tasks. We also highlight the approaches that have shown promising results in semantic segmentation tasks on state-of-the-art datasets. Table 4 highlights some of the most effective methods reported in various semantic segmentation studies.

Table 4. Deep learning methods and network models for semantic segmentation

|  |  |  |
| --- | --- | --- |
| Method | Approach | Sample ConvNet |
| Feature encoder based methods | Learn features by propagating through a series of stacked convolutions layers, activation, and pooling layers. | VGGNet |
| ResNet |
| FRRN |
| ResNeXt |
| Region-based semantic Segmentation | Uses color and similarity across the image to detect regions after which they carry out classification. | R-CNN |
| RPN |
| FPN |
| Mask-RCNN |
| PANet |
| Recurrent neural network-based methods | Utilizes long-term dependencies in sequential data to establish the relationship, sequence, and dependencies within the data.  Commonly applied in time series analysis and change detection. | RNN |
| RNN-CNN |
| ReSeg |
| DAG-RNN |
| DD-RNN |
| 2D-LSTM |
| Directed Acyclic Graph RNNs |
| MCRNNs |
| Up-sampling / de-convolution based methods | Learns high-level features by propagating an image through layers while sacrificing the spatial information. Spatial information lost during sub-sampling and max-pooling operation is regained by upsampling and de-convolution.  Reconstruction technique during upsampling to recover lost spatial resolution for improved classification and semantic segmentation results. | FCN |
| OA-Seg |
| FC-DenseNet |
| SMSNet |
| RefineNet |
| LRN Net |
| DDSC |
| SDN |
| SegNet |
| UNet |
| Increase the resolution of feature-based methods | Recovers the spatial resolution using atrous convolution which generates high-resolution feature maps.  Dilated convolution expands the receptive field by increasing the dilation rate value without losing resolution. | DeepLab v1,2,3, |
| Dilated Convolutions Module |
| SQ Network |
| Hybrid Dilated Convolution (HDC) |
| Dilated Residual Network (DRN) |
| Dense Dilated Spatial Pyramid Pooling (DDSPP) |
| Semi and weakly supervised concept | Uses limited labeled data and abundant unlabeled data to train a deep neural network. | MIL-FCN |
| DecoupledNet |
| BoxSup |
| DecoupledNet |

1. *Feature encoder methods* extract salient features in an image through layer-to-layer propagation thus generating resultant feature maps. VGGNet (Simonyan and Zisserman, 2015) and ResNet (He et al., 2016) are the most dominant approaches that employ feature extraction. Unlike LeNet and AlexNet, VGGNet employs multiple 3 ×3 convolutions in a sequence to match the effect of larger receptive fields, (5 ×5 and 7 ×7). It however requires a large number of parameters and high computational requirements due to large classifiers. ResNet is a very popular network that is widely used for semantic segmentation tasks. It introduced the residual learning concept that enabled networks to attain greater depth and thus learn more features. This ground-breaking concept helped to overcome the vanishing gradient problem associated with plain deep networks. Other networks that use feature encoder methods are FRRN and ResNeXt.
2. *Region-based methods’* main idea is to detect regions in the image according to the similarity of color spaces and other resemblance metrics, and then perform the classification based on the proposed regions that have the highest probability of containing an object. These methods have been greatly employed in computer vision tasks and especially in object detection and can simultaneously achieve both recognition and segmentation (Lateef and Ruichek, 2019). Dissecting regional proposals has been noted to reduce computational power compared to feeding into the network wrapped regions. Region Convolution Neural Networks (R-CNN) can be built on top of any CNN structures. RPN, FPN, Mask-RCNN, and PANet are region-based networks.
3. *Encoder-decoder methods.* CNNs attempts to learn high-level features automatically through layer-to-layer propagation. During this process, the spatial resolution is sacrificed due to the pooling and down-sampling operations. Using the reconstruction method, the features lost in down-sampling are regained through the de-convolution and up-sampling method. Again, low features learnt in the down-sampling are concatenated and fused with high-level features thus improving classification results (Lateef and Ruichek, 2019). Encoder-Decoder methods have been applied in various research works and have achieved competitive results in semantic segmentation tasks(Wang et al., 2018). Following the encoder-decoder approach, the HDC framework was proposed. The framework enlarges the receptive field to aggregate global information better thus improving the standard dilated convolution. Various other studies have used encoder decoder in segmentation for very high-resolution aerial imagery (Diakogiannis et al., 2020), hyperspectral image classification using stacked auto-encoder approach (Kemker and Kanan, 2017), height estimation (Zheng et al., 2019b), and multispectral image segmentation with limited labeled data (Saxena et al., 2020). The encoder-decoder method has recorded more popularity in segmentation tasks compared to other related DL methods.
4. *Conditional Random Fields* - Some architectures have incorporated CRF methods as a post-processing technique to improve segmentation. OA-Seg, FC-DenseNet, SMSNet, RefineNet, LRN-Net, DDSC, and SDN employ the encoder-decoder approach. Challenges on current approaches relating to generalization and stability of the models remain open to further research while demand for signiﬁcantly smaller, more efficient, and faster architectures for mobile and real-time applications remains subject to further research.
5. *Increased resolution feature-based methods.* Most conventional ConvNets produce relatively coarse and low-resolution feature maps due to max-pooling and down-sampling operations which often sacrifice the spatial resolution. This thus requires the low-resolution features to be mapped to input resolution for accurate classification. The mapping function produces high-resolution features required for accurate boundary localization and classification. Increased resolution feature-based methods attempt to preserve the spatial dimensions through atrous convolution (Chen et al., 2017b) also referred to as dilated convolution (Wang et al., 2018) which helps to attain high-resolution feature maps for semantic segmentation tasks. In dilated convolution, the dilation rate defines “the space between the values in a kernel” in the convolution layer. This rate can be expanded without losing spatial resolution. Various studies have employed resolution based methods and achieved promising results in semantic segmentation tasks. DeepLab and its variants, dilated convolutions module, SQ Network, HDC, DRN, and DDSPP are some of the semantic segmentation networks based on increased resolution feature-based methods.
6. *Semi and weakly supervised semantic segmentation* are methods and approaches that have been proposed to overcome the data insufficiency for various image processing tasks. These approaches employ fine-tuning of pre-trained networks for object recognition (Feng and Wang, 2019), where image labels are used to carry out weakly supervised segmentation task. Using this approach, the network learns pixel-level semantic segmentation from weak image labels to indicate the presence or absence of an object where weak annotations are combined with a small number of strong annotations to train DCNN. Networks such as DecoupledNet, BoxSup, and MIL-FCN are some of the networks that have shown remarkable performance in semantic segmentation tasks and object boundary delineation using weak supervision.
   1. *Current trends and perspectives*

*Pre-training and fine-tuning* – as discussed earlier in section , pretraining and fine-tuning have proven very helpful in many computer vision tasks for training DCNNs. Following these approaches, prior trained networks for a given task are adapted to a given target task. Usually, the feature extraction network is pre-trained on a large image classiﬁcation dataset (such as ImageNet and other datasets indicated earlier) and later used in a different dataset with varying class labels and feature distributions (Long et al., 2015a). Also, synthetic data have successfully been used to achieve this objective instead of real data. Generated synthetic data have been used to initialize weights in an end-to-end semantic segmentation network where the network is later fine-tuned with real multispectral data. (Kemker et al., 2018) used Digital Imaging and Remote Sensing Image Generation (DIRSIG) modeling software to generate vast amounts of synthetic Multispectral Images (MSI) and matching label maps. The results show that the approach outperforms traditional classifiers and unsupervised feature extraction techniques. Other studies combined CNN and RNN to pre-train and fine-tune their networks with pseudo labels (Wu and Prasad, 2018), while (Zheng et al., 2019a) employed reinforcement learning to achieve urban segmentation. Following the transfer learning concept, HSI semantic segmentation has been achieved using DL despite limited training data constraints.

*Atrous Convolution* – Whereas atrous convolution (Chen et al., 2017a) also referred to as *dilated convolution* (Yu and Koltun, 2016) is not a new concept in dense prediction, most of the researchers continue to use it for most of the image classification tasks. The method aggregates multi-scale contextual information without losing image resolution. The dilated module consists of convolutional layers with no pooling or subsampling and supports the exponential expansion of the receptive field without loss of resolution or coverage and can be embedded in other architectures. This type of convolution improves object localization with a small receptive field and balances the context assimilation with a larger receptive field which leads to better precision and more detailed segmentation maps. DeepLabV2 (Chen et al., 2017a), proposed an Atrous spatial pyramid pooling (ASPP) for robust segmentation of objects at multiple scales. DeepLabV3 uses cascaded or parallel atrous convolution to capture multi-scale context and outperformed its predecessors. The motivation of atrous convolution is to provide finer and precise object segmentation maps while maintaining spatial resolution which is paramount for delineating segmentation classes.

*Weak Supervision -* Some of the studies that have shown impressive results in semantic segmentation using weak supervision concept are Precision Enhancement Method for Semantic Segmentation Network (PESSN) (Park et al., 2019), which addresses over-segmentation by defining confidence-based and semantic-correlation-based outliers and DecoupledNet (Liu et al., 2018), which employs two separate networks (one for object label classification and another for segmentation) to perform segmentation for close intra-class variations.

*Attention Mechanism -* Recently, most of the DCNNs models for image classification have shifted the paradigm from not only concentration on the development of deep and wide-layered networks, but to network with more robust layers and enhanced feature extraction methods that can fully capture all the semantics necessary in a feature descriptor. Attention mechanism (Vaswani et al., 2017) is one of such methods that aim at learning weight map which represents the relative importance of activations within the same layer or channel. Attention mechanism has been explored in image classification tasks for HSI classification to guide the network in discriminating the essential spatial and spectral channels to generate a sufficient feature descriptor (Hang et al., 2020; Li et al., 2020b; Liu et al., 2020c; Luo et al., 2019a). Attention mechanism continues to draw interest as researchers combine the method with other approaches like atrous convolution and encoder-decoder-based methods (Yang et al., 2020), and the multiscale feature extraction method (Zhang et al., 2020).

*Multi-level feature aggregation* – In this new approach, features are extracted at different spatial dimensions in a multi-leveled and multi-layered structured network. The outputs of the preceding layers are fed into the input of the successive layers resulting in the fusion of sufficient spectral and spatial information where multiscale-feature upsampling blocks are usually used to increase the size of combined feature maps with different resolutions to utilize the information from different sizes and locations. This method has been explored in various hyperspectral tasks such as land cover classification (Cao et al., 2018), spectral fusion, and HSI image classification (Han et al., 2019; Mu et al., 2020). The method has been found efficient in strengthening feature propagation and improving the accuracy of downstream tasks by extracting and fusing multilevel convolutional features from different CNN layers (Zhuo and Wang, 2018).

*Generative adversarial networks (GANs)* – As discussed earlier in section , GANs help address training label insufficiency by learning to generate samples from data distribution using two competing neural networks, namely a generator and a discriminator. Several works (Alipourfard and Arefi, 2020; Tao et al., 2020; Zhu et al., 2018) have all accomplished promising results in hyperspectral image classification using GANs, (Feng et al., 2020) combined GANs with attention mechanism to improve the model’s performance, while (Zhong et al., 2020) enhanced the results in HSI by incorporating a conditional random field (CRF) module to refine the classification results.

1. **Experimental results and discussions**

In this section, we discuss several models that have been proposed for semantic segmentation tasks and have demonstrated compelling results. We include the standard generic datasets following the intuition that most HSI models are first pre-trained on largescale generic datasets and later transfer knowledge to other domains including HSI tasks. This will present the researchers with a general overview for comparison on dataset’s complexity and model’s performance when choosing base network architecture for a given segmentation task. We have also presented results of popular HSI segmentation datasets. We have shared their corresponding approaches and performances as reported in their original reports.

* 1. *Datasets used for evaluation*

Most of the research on image segmentation have used 2D image datasets. In this study, we report 4 of the most popular datasets used by most of the researchers as benchmarks for semantic segmentation tasks.

1. *PASCAL Visual Object Classes (VOC)* (Chen et al., 2014a) is noted as one of the most popular datasets in image processing domains. It contains annotated images with 21 classes with object labels of vehicles, household, animals, airplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, TV/monitor, bird, cat, cow, dog, horse, sheep, and person. Most of the algorithms have extensively used it as a benchmark dataset for various image-processing tasks such as detection, recognition, classification, and segmentation. In this dataset, pixels that do not belong to the stated classes are labeled as background. This dataset is divided into two sets, training, and validation, with 1,464 and 1,449 images.
2. *Microsoft Common Objects* in Context (MS COCO) (Lin et al., 2015) contains 91 generic object types evident in our everyday life with a total of 2.5 million labeled instances of 328k images. It also includes complex image scenes with objects under their rich natural settings, large-scale object detection, segmentation, and captioning dataset. It has been popularly used for segmentation tasks where 82k images are set for training, 40.5k images for validation, and more than 80k images for the test set.
3. *Cityscapes Dataset* (Cordts et al., 2015) is a large-scale database developed for semantic understanding of urban street scenes and urban understanding. It contains stereo video sequences recorded in 50 cities with different street scenes. It contains high-quality pixel-level annotation of 5k frames, with an additional set of 20k weakly annotated frames. It has semantic and dense pixel annotations of 30 classes, grouped into 8 categories—flat surfaces, humans, vehicles, constructions, objects, nature, sky, and void.
4. *ADE20K /MIT* (Zhou et al., 2017) is a standard dataset platform for training and testing Scene Parsing algorithms. This dataset contains more than 20K scene centric generic images well annotated with objects and object parts. The benchmark dataset is divided into 20K images (training), 2K images (validation), and another batch of images for testing. The ADE20K dataset has 150 semantic categories.

We demonstrate the segmentation results for various models on select datasets. To have a standard metric, we chose to report on models whose evaluation scores were based on the mean intersection over union (*mIoU*) metric.

* 1. *Evaluation metrics used in semantic segmentation*

Regular performance evaluation metrics for image segmentation include Intersection over union, pixel accuracy, precision, and recall. All of the four metrics are described below (Long et al., 2015a):

*Intersection Over Union (IoU),* is one of the most commonly used metrics in semantic segmentation. It is defined as the area of intersection between the predicted segmentation map and the ground truth, divided by the area of union between the predicted segmentation map and the ground truth:

IoU = J(A, B) = |A ∩ B| / |A ∪ B| (1)

Where A and B denote the ground truth and the predicted segmentation maps, respectively. It ranges between 0 and 1.

The IoU score is calculated for each class separately and then averaged over all classes to provide a global, mean IoU score of our semantic segmentation prediction.

Pixel accuracy indicates the percent of pixels in the image which are correctly classified. The pixel accuracy is usually reported for each class separately as well as globally across all classes. Pixel accuracy is defined as:

(2)

Precision describes the clarity of our positive detections as compared to the ground truth. Precision considers all the objects predicted in a given image and computes how many of those objects have a matching ground truth annotation.

(3)

This metric can sometimes provide misleading results when the class representation is small within the image.

Recall describes the completeness of our positive predictions relative to the ground truth. It measures all positive predictions relative to all of the objected annotated in our ground truth.

(4)

* 1. *Loss functions used in semantic segmentation*

We discuss the common loss functions used in pixel-level image classification. More information on the loss function discussed here can be found in the works of Jadon et al., (Jadon, 2020)

Cross-categorical entropy loss, also referred to as logistic loss and multinomial logistic loss is defined as:

(5)

Where ti and si are the ground-truth and the model score for each class i in C. An activation function is applied to the scores before the cross-entropy loss is computed, we write f (Si) to refer to the activations.

*Focal loss* is an improvement to the standard cross-entropy criterion. It is computed by changing its shape such that the loss assigned to well-classified examples is down-weighted. Eventually, this ensures that there is no class imbalance. In this loss function, the cross-entropy loss is scaled with the scaling factors decaying at zero as the confidence in the correct classes increases. The scaling factor automatically down weights the contribution of easy examples at training time and focuses on the hard ones.

FL(pt) = -(1 - pt)γ log(pt). (6)

*Dice loss* is obtained by calculating the smooth dice coefficient function. This loss is the most commonly used loss function for segmentation problems (Huang et al., 2020).

(7) *Intersection over Union balanced loss* aims at increasing the gradient of samples with high IoU and decreasing the gradient of samples with low IoU thus increasing the localization accuracy of machine learning models.

*IoU = TP / (TP + FP + FN)* (8)

*Boundary loss* is applied to tasks with highly unbalanced segmentations. This loss’s form is that of a distance metric on space contours and not regions. In this manner, it tackles the problem posed by regional losses for highly imbalanced segmentation tasks.

(9)

*Weighted cross-entropy loss* - This loss is used in scenarios that involve class imbalance where all positive examples are weighted by a certain coefficient.

(10)

*Lovász-Softmax loss* - This loss performs direct optimization of the mean intersection-over-union loss in neural networks based on the convex Lovasz extension of sub-modular losses.

(11)

Since image segmentation task is not only concerned with the location of the pixel at the local context but also considers the relationship between a given pixel with the rest at the global context, using different loss functions for segmentation task have shown promising results in recent works.

* 1. *Results and discussions*

The following *Table 5*, *Table 6*, *Table 7,* and *Table 8* presents the performance of previously discussed algorithms on popular segmentation dataset benchmarks. We have considered PASCAL VOC, MS COCO, ADE20K, and Cityscapes in this study as the generic benchmark datasets for comparing the selected models. Using generic datasets for comparison informs the researcher on the dataset’s complexity as well as the general model’s performance on standard image dataset which can guide the choice of baseline network architecture in case a researcher opts to use transfer learning in HSI segmentation task. As stated earlier, generic image datasets are important for pre-training image classification models that are later fine-tuned for HSI image segmentation tasks.

Table 5: Segmentation models results based on the PASCAL VOC test set

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Baseline  Network | Metric  (mIoU) | Year |
| FCN (Long et al., 2015a) | VGG-16 | 62.2 | 2015 |
| CRF-RNN (Zheng et al., 2015) | undefined | 72 | 2015 |
| CRF-RNN (Zheng et al., 2015) | undefined | 74.7 | 2015 |
| BoxSup (Dai et al., 2015) | undefined | 75.1 | 2015 |
| DPN (Liu et al., 2015) | undefined | 77.5 | 2015 |
| DeepLab-CRF (Chen et al., 2017a) | ResNet-101 | 79.7 | 2018 |
| GCN (Peng et al., 2017) | ResNet-152 | 82.2 | 2017 |
| RefineNet(Lin et al., 2017b) | ResNet-152 | 84.2 | 2017 |
| Wide ResNet(Wu et al., 2019) | WideResNet-38 | 84.9 | 2018 |
| PSPNet (Zhao et al., 2017) | ResNet-101 | 85.4 | 2017 |
| DeeplabV3 (Chen et al., 2017b) | ResNet-101 | 85.7 | 2017 |
| PSANet (Zhao et al., 2018) | ResNet-101 | 85.7 | 2018 |
| DFN (Yu et al., 2018) | ResNet-101 | 82.7 | 2018 |
| Exfuse (Zhang et al., 2018) | ResNet-101 | 86.2 | 2018 |
| SDN (Fu et al., 2017b) | DenseNet-161 | 83.5 | 2017 |
| SDN (Fu et al., 2017b) | DenseNet-161 | 86.6 | 2017 |
| DM-Net (He et al., 2019a) | ResNet-101 | 84.4 | 2019 |
| DM-Net (He et al., 2019a) | ResNet-101 | 87.06 | 2019 |
| APC-Net (He et al., 2019b) | ResNet-101 | 84.2 | 2019 |
| APC-Net (He et al., 2019b) | ResNet-101 | 87.1 | 2019 |
| DeeplabV3+ (Chen et al., 2018) | Xception-71 | 87.8 | 2018 |
| Exfuse (Zhang et al., 2018) | ResNeXt 131 | 87.9 | 2018 |
| MSCI (Lin et al., 2018a) | ResNet 152 | 88 | 2018 |
| EMANet (Yi et al., 2019) | ResNet 152 | 88.2 | 2019 |
| DeeplabV3+ (Chen et al., 2018) | Xception 71 | 89 | 2018 |

Table 6. Semantic segmentation models on the Cityscapes dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Baseline | Metric (mIoU) | Year |
| DeeplabV2 CRF (Chen et al., 2017a) | ResNet 101 | 70.4 | 2017 |
| RefineNet (Lin et al., 2017b) | ResNet 101 | 73.6 | 2017 |
| FoveaNet (Li et al., 2017) | ResNet 101 | 74.1 | 2017 |
| Ladder (Krapac and Šegvic, 2017) | DenseNet 169 | 73.7 | 2017 |
| GCN (Peng et al., 2017) | ResNet 101 | 76.9 | 2017 |
| DUC-HDC (Wang et al., 2018) | ResNet 101 | 77.6 | 2017 |
| Wide ResNet (Wu et al., 2019) | WideResNet 38 | 78.4 | 2019 |
| Dense ASPP (Yang et al., 2018) | Densenet 161 | 80.6 | 2017 |
| DeepLabV3 (Chen et al., 2017b) | ResNet-101 | 81.3 | 2018 |
| DeepLabV3+ (Chen et al., 2018) | Xception-71 | 82.1 | 2018 |
| GS-CNN (Takikawa et al., 2019) | WideResNet | 82.8 | 2019 |

*Table* 5 focuses on the results of various models on the PASCAL VOC test set. The introduction of FCN contributed to a remarkable improvement in the accuracy of the models and remains the most popular DL-based image segmentation base concept. The latest DeepLabV3+ feature about 27.3% points margin relative gain over the initial FCN model on the PASCAL VOC dataset. Table 6 focuses on the Cityscape test dataset while Table 7 focuses on the MS COCO stuff test set. Some models have been pre-trained on the ImageNet dataset while most of the models are fully trained on a full network (FN).

This dataset is more challenging than PASCAL VOC, and Cityscapes, as the highest mIoU is approximately 40%. Table 8 focuses on the ADE20k validation set. This dataset is also more challenging than the PASCAL VOC and Cityscapes datasets.

Table 7. Segmentation model’s results based on the MS COCO stuff dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Baseline  Network | Metric  (mIoU) | Year |
| RefineNet (Lin et al., 2017b) | ResNet 101 | 33.6 | 2017 |
| CCN (Ding et al., 2018) | Ladder DenseNet 101 | 35.7 | 2018 |
| DSSPN (Liang et al., 2018) | ResNet 101 | 37.3 | 2018 |
| EMA-Net (Li et al., 2019b) | ResNet 50 | 37.5 | 2019 |
| OCR (Yuan et al., 2019) | ResNet 101 | 39.5 | 2019 |
| EMA-Net (Li et al., 2019b) | ResNet 50 | 39.9 | 2019 |
| OCR (Yuan et al., 2019) | HRNetV2 W48 | 40.5 | 2019 |

Table 8. Segmentation model’s results based on the ADE20k validation dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Baseline  Network | Metric  (mIoU)(%) | Year |
| FCN (Long et al., 2015a) | undefined | 29.39 | 2015 |
| RefineNet (Lin et al., 2017b) | ResNet 152 | 40.7 | 2017 |
| PSPNet (Zhao et al., 2017) | ResNet 101 | 43.29 | 2017 |
| SAC (Zhang et al., 2017) | ResNet 101 | 44.3 | 2017 |
| UperNet (Xiao et al., 2018) | ResNet 101 | 42.66 | 2018 |
| DSSPN (Liang et al., 2018) | ResNet 101 | 43.68 | 2018 |
| PSANet (Zhao et al., 2018) | ResNet 101 | 43.7 | 2018 |
| PSPNet (Zhao et al., 2017) | ResNet 269 | 44.94 | 2017 |
| OCR (Yuan et al., 2019) | HRNetV2 W48 | 45.6 | 2019 |
| AC-Net (Fu et al., 2019) | ResNet 101 | 45.9 | 2019 |

Table 9. Performance of various DCNNs models on HSI datasets

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Testing Dataset | Method | Score / Metrics |
| DDCM-Net (Liu et al., 2020b) | DeepGlobe | Dilated convolution | 56.2% mIoU |
| DiResNet (Ding and Bruzzone, 2020) | DeepGlobe | Residual connections and deconvolution | 98.44% OA |
| ATD-LinkNet (Qi et al., 2020) | DeepGlobe | Multiscale and attention mechanism | 62.68% mIoU |
| DLinkNet (Zhou et al., 2018) | DeepGlobe | Dilated convolution | 63.42% mIoU |
| WGAN-GP (Yang and Wang, 2020) | DeepGlobe | GAN | 73.0% mIoU |
| DeepLabv3+ (Chen et al., 2018) | ISPRS (PotsDam) | Dilated convolution | 66.82% mIoU |
| ATD-LinkNet (Qi et al., 2020) | ISPRS (PotsDam) | Multiscale and attention mechanism | 89% OA |
| SCAttNet (Li et al., 2020a) | ISPRS (PotsDam dataset) | Channel and Spatial attention mechanism | 68.31% mIoU |
| SegNet (Badrinarayanan et al., 2017) | ISPRS (Vaihingen dataset) | Encoder-decoder | 64.06% mIoU |
| RefineNet (Lin et al., 2017a) | ISPRS (Vaihingen dataset) | Multiple paths | 69.19% mIoU |
| DenseU-net (Dong et al., 2019) | ISPRS (Vaihingen dataset) | Encoder-decoder | 85.63% OA |
| ResUNet (Diakogiannis et al., 2020) | ISPRS 2D Potsdam dataset. | Residual connections | 92.9% OA |
| Denoising autoencoders (Nalepa et al., 2020a) | Salinas scene | Autoencoders | 98.54% OA |
| Graph-based spatial-spectral feature learning (Ahmad et al., 2017) | Salina | Graph-based method | 97.83% OA |
| Resource-frugal quantized convolutional neural networks (Nalepa et al., 2020a) | Pavia Center | Quantized convolutions | 91.73% OA |
| Denoising autoencoders (Nalepa et al., 2020a) | Pavia Center | Autoencoders | 96.96% OA |
| Graph-based spatial-spectral feature learning (Ahmad et al., 2017) | Pavia Center | Graph-based method | 98.54% OA |
|  |  |  |  |
| SCAE (Kemker and Kanan, 2017) | Indian Pines | Self-taught learning | 98.03% OA |
| 2D–3D CNN (Ge et al., 2020) | Indian Pines | Multibranch feature fusion | 96.07% OA |
| Graph-based spatial-spectral feature learning (Ahmad et al., 2017) | Indian Pines | Graph-based method | 98.38% OA |
| PL-SSDL (Wu and Prasad, 2018) | University of Houston | A semi-supervised method with pseudo labels | 82.61% OA |
| PL-SSDL (Wu and Prasad, 2018) | University of Pavia dataset | A semi-supervised method with pseudo labels | 88.43% OA |
| PL-SSDL (Wu and Prasad, 2018) | Wetland dataset | Semi-supervised method with pseudo labels | 97.33% OA |
| Pop-Net (Zheng et al., 2019b) | IEEE GRSS Data Fusion dataset | Encoder with Dual decoder | 77.78% OA |

*Table 9* presents the common HSI datasets used in semantic segmentation tasks together with the results of the current models on these datasets as reported in their original papers.

In the reported models in Table 9, GAN based network (WGAN-GP) demonstrates better performance over dilated (DDCM-Net) and multi-scale methods using the attention mechanism (ATD-LinkNet) on the DeepGlobe dataset with a leading mIoU of 73.0%. In the ISPRS Potsdam dataset, the network model using the Attention mechanism (SCAttNet) reports better *mIoU* scores of 68.31% than the dilated convolution method (DeepLabV3+) which scored 66.82%. Also, RefineNet based on multiple paths performed better with a 5% score over SegNet which is based on Encoder-Decoder. Autoencoders have performed better than graph-based methods in the Salina dataset but on the other hand, scored lower than graph-based methods in the Pavia center dataset. This can be attributed to the difference in dataset complexity and network parameter settings. In the Indian Pines dataset, the Graph-based method reported better overall accuracy compared to the multi-branch feature fusion method and self-taught learning approach.

Whereas most models in various reports have used standard metrics, some models have not tested their results on standard benchmarks thus making it difficult to compare the performances on level ground. In addition, only a small percentage of the works have provided additional information on computational resources, (execution time, and memory usage) which are important parameters for comparison on the performances of these models.

Most of the current tasks on image segmentation using the HSI dataset still use well-known methods such as FCN, residual connections, encoder-decoder, dilated convolutions, and semi-supervised methods with pseudo labels. In the recent literature, new techniques such as graph-based methods, attention mechanism, multi-layer and multi-branch feature fusion and GANs are being adopted for higher performance and precision in the feature extraction process. We summarize these trends on most of the HSI segmentation tasks in the next section.

1. **Discussions on current trends.**

FCN sets the foundation of most of the current well-performing segmentation models due to its distinguished contribution to overcoming the fully connected network input size limitation. UNet was initially developed for bio-medical image segmentation is based on the encoder-decoder paradigm. It has been applied in many image segmentation tasks and is rated as one of the most popular image segmentation networks. It solves the information loss problem by use of skip connections to send information to every upsampling layer in the decoder from the corresponding downsampling layer in the encoder*.*

Atrous convolution has been used in many models to probe image features without increasing the number of parameters. It works by appending holes to the kernel to increase the size of the filter and have been applied in many studies for capturing large context in objects. Spatial Pyramidal Pooling (ASPP) concept has been applied in many networks to achieve multiscale information from the feature map. Instead of using a single image, the ASPP concept replaces the need to have different images of different sizes to aggregate multiscale context.

Cascaded layers treat a single deep model as a cascade of several sub-models and classiﬁes most of the easy regions in the shallow stage and make the deeper stage focus on a few hard regions. By introducing different levels in the network, the layer cascade model can probe coarse and fine details as different resolutions thus improving accuracy and speed. Most DCNNs based HSI classification networks are based on either encoder or fully convolution network and incorporate other methods discussed earlier as part of their network to capture the benefits of both precise localization and to capture robust features descriptor.

Current trends on image segmentation have shifted the focus on how best to extract a full representation of feature descriptors while maintaining low computation complexities. Various researchers have used different layer levels to extract both spatial and aspatial features, as others included attention mechanisms to guide the network on what features to extract and capture. Besides, some works have employed multi-level feature fusion where features are extracted at different spatial dimensions in a multi-leveled and multi-layered structured network. GANs have also gained wide acceptance in HSI classification tasks with some researchers incorporating attention mechanisms in most networks to refine the precision and robustness of their models. Lastly, models that have combined different approaches have shown better performances in classification results despite recording some increased computation load. We, however note that these methods have also been reported effective in other image processing tasks such as medical image segmentation, and object detection, classification of generic images among others (Signoroni et al., 2019a)

Conclusion

The findings of this study can be understood as the performance of deep learning hyper spectral imaging tasks is lagging as compared to object identification and scenario recognition. The generalized datasets used in transfer learning do not represent diverse objects enough to represent the real HSI world where they exhibit inherent characteristics such as shadows, varied lightning conditions, occlusion and overlapping shades. As some approaches seem better in tackling data insufficiency challenges in HSI segmentations. Some researchers alternatively suggested the generation of synthetic datasets, adopting data augmentation techniques, transfer learning, domain adaptation, unsupervised function learning and network optimization methodologies.

The literature has laid emphasis on incorporating layer level optimization methods in their models and seek to leverage adequate features from the limited labeled dataset for optimal descriptors to attain better scores on efficiency, reduce computation costs, and develop more dynamic models. In the coming years, we expect that more intuitive techniques and methods can be used to solve key challenges in the dynamic classification of HIS images and to navigate different remote sensing datasets to unlock the full potential of HIS applications. This literature review presented various accessible datasets including synthetic datasets and outlined the promising approaches for HIS image processing tasks to handle data insufficiency. Lastly, we have concluded the basis of the current methods and future trends in hyperspectral imaging research.

Acknowledgments

This work was supported by the Fujian key Laboratory for Sensing and Computing for Smart Cities of Fujian Province of China.

Abbreviations

The following abbreviations are used in this manuscript

AC-Net Adaptive context network

ANN Artificial neural networks

ASPP Atrous spatial pyramid pooling

CRF Conditional random fields

CRNN Convolution recurrent neural network

DA Domain adaptation

DANet Dual attention network

DBN Deep belief network

DCNN Deep convolution neural network

DDSPP Dense dilated spatial pyramid pooling

DIRSIG Digital imaging and remote sensing image generation

DL Deep learning

DM-NET Adaptive pyramid context network

DPN Deep parsing network

DUC-HDC Dense up-sampling convolution – hybrid dilated convolution

EMA-Net Expectation-maximization attention networks

ESA European space agency

FCN Fully convolutional neural network

GAN Generative adversarial networks

GCN Global convolutional network

GS-CNN Gated shape CNNs

HSI Hyperspectral imagery

HSI-RS High spatial resolution remote sensing

ISPRS International Society for photogrammetry and remote sensing

MLP Multilayer perceptron

MSCI Multi-scale context intertwining

MSI Multispectral images

NASA National aeronautics and space administration.

NCALM National center for airborne laser mapping

OCR Object-contextual representations

PL-SSDL semi-supervised deep learning using pseudo labels

PSPNet Pyramid scene-parsing network

R-CNN Region convolution neural networks

ReLU Rectified linear unit

RF Random forest

SAC Scale-adaptive convolutions

SAE Stacked auto-encoders

SAR Synthetic aperture radar

SDN Stacked deconvolution network

SVM Support vector machines

UDA Universal domain adaptation

**References**

Ahmad, E., Goyal, M., Mcphee, J., Degens, H., 2018. Semantic Segmentation of Human Thigh Quadriceps Muscle in Magnetic Resonance Images. arXiv 1.

Ahmad, M., Khan, A.M., Hussain, R., 2017. Graph-based spatial–spectral feature learning for hyperspectral image classification. IET Image Processing 11, 1310-1316.

Alan, J.B., Tateishi, R., Hoan, N.T., 2013. A hybrid pansharpening approach and multiscale object-based image analysis for mapping diseased pine and oak trees. International journal of remote sensing 34.20, 6969–6982.

Alipourfard, T., Arefi, H., 2020. Structure Aware Generative Adversarial Networks for Hyperspectral Image Classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 5424-5438.

Alvarez, J.M., Gevers, T., LeCun, Y., Lopez, A.M., 2012. Road Scene Segmentation from a Single Image, in: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (Eds.), Computer Vision – ECCV 2012. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 376-389.

Angus, M., ElBalkini, M., Khan, S., Harakeh, A., Andrienko, O., Reading, C., Waslander, S., Czarnecki, K., 2018. Unlimited Road-scene Synthetic Annotation (URSA) Dataset. arXiv.

Aptoula, E., Ozdemir, M.C., Yanikoglu, B., 2016. Deep Learning With Attribute Profiles for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters PP, 1-5.

Armeni, I., Sax, S., Zamir, A.R., Savarese, S., 2017. Joint 2D-3D-Semantic Data for Indoor Scene Understanding. arXiv:1702.01105 2.

Ba, J.L., Kiros, J.R., Hinton, G.E., 2016. Layer Normalization. arXiv:1607.0645 1.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 2481-2495.

Bahmanyar, R., Vig, E., Reinartz, P., 2019. MRCNet: Crowd Counting and Density Map Estimation in Aerial and Ground Imagery.

Ball, J., Anderson, D., Chan, C.S., 2017. A Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools and Challenges for the Community. Journal of Applied Remote Sensing 11.

Ballan, L., Taneja, A., Gall, J.u., Gool, L.V., Pollefeys, M., 2012. Motion Capture of Hands in Action using Discriminative Salient Points. ECCV.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., Nemani, R., 2015. DeepSat – A Learning framework for Satellite Imagery. Association for Computing Machinery.

Battaglia, P.W., Hamrick, J.B., Tenenbaum, J.B., 2013. Simulation as an engine of physical scene understanding. Proc Natl Acad Sci U S A 110, 18327-18332.

Bellman, R.E., 2015. Adaptive control processes: a guided tour. Princeton university press.

Benjdira, B., Bazi, Y., Koubaa, A., Ouni, K., 2019. Unsupervised Domain Adaptation Using Generative Adversarial Networks for Semantic Segmentation of Aerial Images. Remote Sens. 11.

Bhoyar, K., 2014. Synthetic Image Dataset.

Bousmalis, K., Irpan, A., Wohlhart, P., Bai, Y., Kelcey, M., Kalakrishnan, M., Downs, L., Ibarz, J., Pastor, P., Konolige, K., Levine, S., Vanhoucke, V., 2018. Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping.

Brian, J., Tateishi, Xie, R., 2012. Using geographically-weighted variables for image classification. Remote Sensing Letters.

Brostow, G.J., Fauqueur, J., Cipolla, R., 2009. Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters 30, 88-97.

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A., 2020. Albumentations: Fast and Flexible Image Augmentations. Inf. 11, 125.

Caesar, H., Uijlings, J.R.R., Ferrari, V., 2018. COCO-Stuff: Thing and Stuff Classes in Context. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1209-1218.

Cao, X., Li, R., Wen, L., Feng, J., Jiao, L., 2018. Deep Multiple Feature Fusion for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 3880-3891.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. arXiv 1.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017a. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. arXiv 2.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017b. Rethinking Atrous Convolution for Semantic Image Segmentation.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, CVPR. Springer.

Chen, X., Mottaghi, R., Liu, X., Fidler, S., Urtasun, R., Yuille, A., 2014a. Detect What You Can: Detecting and Representing Objects Using Holistic Models and Body Parts, 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1979-1986.

Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P., 2016a. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing 54, 6232-6251.

Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P.J.I.T.o.G., Sensing, R., 2016b. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. 54, 6232-6251.

Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014b. Deep Learning-Based Classification of Hyperspectral Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 2094-2107.

Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y.J.I.J.o.S.t.i.a.e.o., sensing, r., 2014c. Deep learning-based classification of hyperspectral data. 7, 2094-2107.

Chen, Y., Zhao, X., Jia, X., 2015a. Spectral–Spatial Classification of Hyperspectral Data Based on Deep Belief Network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8, 2381-2392.

Chen, Y., Zhao, X., Jia, X.J.I.J.o.S.T.i.A.E.O., Sensing, R., 2015b. Spectral–spatial classification of hyperspectral data based on deep belief network. 8, 2381-2392.

Cheng, G., Han, J., Lu, X., 2017. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE 105, 1865-1883.

Cordts, M., Omran, M., Benenson, R., Ramos, S., Franke, U., Scharwachter, T., Enzweiler, M., Roth, S., Schiele, B., 2015. The Cityscapes Dataset. CVPR.

Dai, A., Chang, A., Savva, M., Halber, M., Funkhouser, T., NieBner, M., 2017. ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes.

Dai, J., He, K., Sun, J., 2015. BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation, 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1635-1643.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R.J.I.C.C.o.C.V., Workshops, P.R., 2018. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. 172-17209.

Deng, C., Xue, Y., Liu, X., Li, C., Tao, D., 2019. Active Transfer Learning Network: A Unified Deep Joint Spectral–Spatial Feature Learning Model for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing 57, 1741-1754.

Deng, J., Dong, W., Socher, R., Li, L., Kai, L., Li, F.-F., 2009. ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255.

Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing 162, 94-114.

Ding, H., Jiang, X., Shuai, B., Liu, A.Q., Wang, G., 2018. Context Contrasted Feature and Gated Multi-scale Aggregation for Scene Segmentation, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2393-2402.

Ding, L., Bruzzone, L., 2020. DiResNet: Direction-Aware Residual Network for Road Extraction in VHR Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing, 1-12.

Dong, R., Pan, X., Li, F., 2019. DenseU-Net-Based Semantic Segmentation of Small Objects in Urban Remote Sensing Images. IEEE Access 7, 65347-65356.

Doon, R., Rawat, T., Gautam, S., 2018. Cifar-10 Classification using Deep Convolutional Neural Network.

Dosovitskiy, A., Fischer, P., Ilg, E., Häusser, P., Hazirbas, C., Golkov, V., Smagt, P.v.d., Cremers, D., Brox, T., 2015. FlowNet: Learning Optical Flow with Convolutional Networks, 2015 IEEE International Conference on Computer Vision (ICCV), pp. 2758-2766.

Dua, Y., Kumar, V., Singh, R.S., 2020. Comprehensive review of hyperspectral image compression algorithms. 59 %J Optical Engineering, 090902.

Fan, W., Wei, T., Peng, J., Sun, W., 2019. Ideal Regularized Kernel Subspace Alignment for Unsupervised Domain Adaptation in Hyperspectral Image Classification, 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), pp. 1-4.

Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R.J.I.T.o.G., Sensing, R., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. 46, 3804-3814.

Feng, J., Feng, X., Chen, J., Cao, X., Zhang, X., Jiao, L., Yu, T., 2020. Generative Adversarial Networks Based on Collaborative Learning and Attention Mechanism for Hyperspectral Image Classification. Remote. Sens. 12, 1149.

Feng, Y., Wang, L., 2019. A Weakly-Supervised Approach for Semantic Segmentation, 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 2311-2314.

Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017a. Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. Remote Sensing 9.

Fu, J., Liu, J., Wang, Y., Li, Y., Bao, Y., Tang, J., Lu, H., 2019. Adaptive Context Network for Scene Parsing, 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6747-6756.

Fu, J., Liu, J., Wang, Y., Lu, H., 2017b. Stacked Deconvolutional Network for Semantic Segmentation. IEEE Transactions on Image Processing PP.

Gaetano, R., Masi, G., Poggi, G., Verdoliva, L., Scarpa, G., 2015. Marker-Controlled Watershed-Based Segmentation of Multiresolution Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 53, 2987-3004.

Gaidon, A., Wang, Q., Cabon, Y., Vig, E., 2016. VirtualWorlds as Proxy for Multi-object Tracking Analysis, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4340-4349.

Gao, L., Li, J., Khodadadzadeh, M., Plaza, A., Zhang, B., He, Z., Yan, H.J.I.G., Letters, R.S., 2014. Subspace-based support vector machines for hyperspectral image classification. 12, 349-353.

Ge, Z., Cao, G., Li, X., Fu, P., 2020. Hyperspectral Image Classification Method Based on 2D-3D CNN and Multibranch Feature Fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 5776-5788.

Gerke, M., Rottensteiner, F., Wegner, J., Sohn, G., 2014. ISPRS Semantic Labeling Contest.

Gould, S., Fulton, R., Koller, D., 2009. Decomposing a scene into geometric and semantically consistent regions, 2009 IEEE 12th International Conference on Computer Vision, pp. 1-8.

Graña, M., Veganzons, M., Ayerdi, B., 2013-2018. Hyperspectral Remote Sensing Scenes, in: group, C.I. (Ed.), Faculty of Informatics of San Sebastián, Donostia (Gipuzkoa - Spain).

Grauman, Shakhnarovich, Darrell, 2003. Inferring 3D structure with a statistical image-based shape model, Proceedings Ninth IEEE International Conference on Computer Vision, pp. 641-647 vol.641.

Gupta, A., Vedaldi, A., Zisserman, A., 2016. Synthetic Data for Text Localisation in Natural Images, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2315-2324.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J., Schindler, K., Pollefeys, M., 2017. Semantic3D.net: A new Large-scale Point Cloud Classification Benchmark. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences IV-1/W1.

Hackel, T., Wegner, J., Savinov, N., Ladicky, L., Schindler, K., Pollefeys, M., 2018. Large-Scale Supervised Learning For 3D Point Cloud Labeling: Semantic3d.Net. Photogrammetric Engineering and Remote Sensing 84, 297-308.

Han, X.-H., Zheng, Y., Chen, Y.-W., 2019. Multi-Level and Multi-Scale Spatial and Spectral Fusion CNN for Hyperspectral Image Super-Resolution, pp. 4330-4339.

Hang, R., Li, Z., Liu, Q., Ghamisi, P., Bhattacharyya, S.S., 2020. Hyperspectral Image Classification with Attention Aided CNNs. CoRR abs/2005.11977.

He, J., Deng, Z., Qiao, Y., 2019a. Dynamic Multi-Scale Filters for Semantic Segmentation, 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3561-3571.

He, J., Deng, Z., Zhou, L., Wang, Y., Qiao, Y., 2019b. Adaptive Pyramid Context Network for Semantic Segmentation, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7511-7520.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.

Hoffman, J., Guadarrama, S., Tzeng, E., Donahue, J., Girshick, R., Darrell, T., Saenko, K., 2014. LSDA: Large Scale Detection Through Adaptation. arXiv:1407.5035 2.

Hoffman, J., Rodner, E., Darrell, T., Donahue, J., Saenko, K., 2013. Efficient Learning of Domain-invariant Image Representations. arXiv:1301.3224 5.

Hu, G., Peng, X., Yang, Y., Hospedales, T.M., Verbeek, J., 2018. Frankenstein: Learning Deep Face Representations Using Small Data. IEEE Transactions on Image Processing 27, 293-303.

Huang, B., Zhao, B., Song, Y., 2018a. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sensing of Environment 214, 73-86.

Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y., Yang, R., 2018b. The ApolloScape Dataset for Autonomous Driving. CVPR.

Huang, Y., Tang, Z., Su, K., 2020. Balance-batch: An Optimized Method for Semantic Segmentation Loss Functions, 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), pp. 342-346.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv: 1502.03167, 448–456.

J, B.D., J, W., B, S.G., J, B.M., 2012. A naturalistic open source movie for optical flow evaluation, in: al., A.F.e. (Ed.), European Conf. on Computer Vision. Springer-Verlag, pp. 611--625.

Jadon, S.J.a.p.a., 2020. A survey of loss functions for semantic segmentation.

Jain, S.D., Grauman, K., 2014. Supervoxel-consistent foreground propagation in video, European Conference on Computer Vision. Springer, pp. 656–671.

Kang, X., Li, C., Li, S., Lin, H., 2018. Classification of Hyperspectral Images by Gabor Filtering Based Deep Network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 1166-1178.

Kemker, R., Kanan, C., 2017. Self-Taught Feature Learning for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing 55, 2693-2705.

Kemker, R., Salvaggio, C., Kanan, C., 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing 145, 60-77.

Keysers, D., 2007. Comparison and Combination of State-of-the-art Techniques for Handwritten Character Recognition: Topping the MNIST Benchmark.

Khan, M.J., Khan, H.S., Yousaf, A., Khurshid, K., Abbas, A.J.I.A., 2018. Modern trends in hyperspectral image analysis: a review. 6, 14118-14129.

Kouw, W.M., Loog, M., 2019. TECHNICAL REPORT An introduction to domain adaptation and transfer learning, arXiv:1812.11806, Delft University of Technology, p. 42.

Krapac, J., Šegvic, I.K.S., 2017. Ladder-Style DenseNets for Semantic Segmentation of Large Natural Images, 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), pp. 238-245.

Krizhevsky, A., Sutskever, I., Hinton, G., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems 25.

Kulis, B., Saenko, K., Darrell, T., 2011. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms, CVPR 2011, pp. 1785-1792.

Lai, K., Bo, L., Ren, X., Fox, D., 2011. A large-scale hierarchical multi-view RGB-D object dataset, 2011 IEEE International Conference on Robotics and Automation, pp. 1817-1824.

Laliberte, A.S., Goforth, M.A., Steele, C.M., Rango, A., 2011. Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments. Remote Sensing 3, 2529-2551.

Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., Bulatov, Y., McCord, B.J.A., 2018. xView: Objects in Context in Overhead Imagery. abs/1802.07856.

Lateef, F., Ruichek, Y., 2019. Survey on semantic segmentation using deep learning techniques. Neurocomputing 338, 321-348.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436-444.

Lee, H., Kwon, H., 2017. Going Deeper With Contextual CNN for Hyperspectral Image Classification. IEEE Transactions on Image Processing 26, 4843-4855.

Li, F., Kim, T., Humayun, A., Tsai, D., Rehg, J., 2013a. Video Segmentation by Tracking Many Figure-Ground Segments.

Li, H., Qiu, K., Chen, L., Mei, X., Hong, L., Tao, C., 2020a. SCAttNet: Semantic Segmentation Network With Spatial and Channel Attention Mechanism for High-Resolution Remote Sensing Images. IEEE Geoscience and Remote Sensing Letters, 1-5.

Li, R., Zheng, S., Duan, C., Yang, Y., Wang, X., 2020b. Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network. Remote. Sens. 12, 582.

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., Benediktsson, J.A.J.I.T.o.G., Sensing, R., 2019a. Deep learning for hyperspectral image classification: An overview. 57, 6690-6709.

Li, W., Chen, C., Zhang, M., Li, H., Du, Q., 2018. Data Augmentation for Hyperspectral Image Classification With Deep CNN. IEEE Geoscience and Remote Sensing Letters PP, 1-5.

Li, W., Duan, L., Xu, D., Tsang, I.W., 2014. Learning With Augmented Features for Supervised and Semi-Supervised Heterogeneous Domain Adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence 36, 1134-1148.

Li, W., Prasad, S., Fowler, J.E., Bruce, L.M.J.I.T.o.G., Sensing, R., 2011. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. 50, 1185-1198.

Li, W., Prasad, S., Fowler, J.E.J.I.G., Letters, R.S., 2013b. Hyperspectral image classification using Gaussian mixture models and Markov random fields. 11, 153-157.

Li, W., Tramel, E.W., Prasad, S., Fowler, J.E.J.I.T.o.G., Sensing, R., 2013c. Nearest regularized subspace for hyperspectral classification. 52, 477-489.

Li, X., Jie, Z., Wang, W., Liu, C., Yang, J., Shen, X., Lin, Z., Chen, Q., Yan, S., Feng, J., 2017. FoveaNet: Perspective-aware Urban Scene Parsing.

Li, X., Zhong, Z., Wu, J., Yang, Y., Lin, Z., Liu, H., 2019b. Expectation-Maximization Attention Networks for Semantic Segmentation.

Li, Y., Wang, N., Shi, J., Liu, J., Hou, X., 2016. Revisiting Batch Normalization For Practical Domain Adaptation. arXiv:1603.04779 3.

Liang, X., Zhou, H., Xing, E., 2018. Dynamic-structured Semantic Propagation Network.

Lin, D., Ji, Y., Lischinski, D., Cohen-Or, D., Huang, H., 2018a. Multi-scale Context Intertwining for Semantic Segmentation, ECCV.

Lin, G., Milan, A., Shen, C., Reid, I., 2017a. RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5168-5177.

Lin, G., Milan, A., Shen, C., Reid, I., 2017b. RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation.

Lin, J., He, C., Wang, Z.J., Li, S., 2017c. Structure Preserving Transfer Learning for Unsupervised Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters 14, 1656-1660.

Lin, J., Zhao, L., Li, S., Ward, R., Wang, Z.J., 2018b. Active-Learning-Incorporated Deep Transfer Learning for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 4048-4062.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollar, P., 2015. Microsoft COCO: Common Objects in Context. arXiv:1405.0312 3.

Lin, Z., Chen, Y., Ghamisi, P., Benediktsson, J., 2018c. Generative Adversarial Networks for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing PP.

Liu, B., Fan, H., 2013. Semantic labeling of indoor scenes from RGB-D images with discriminative learning. SPIE.

Liu, C., Yuen, J., Torralba, A., 2009. Nonparametric scene parsing: Label transfer via dense scene alignment, 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1972-1979.

Liu, P., Zhang, H., Eom, K.B., 2017. Active Deep Learning for Classification of Hyperspectral Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 712-724.

Liu, Q., Kampffmeyer, M., Jenssen, R., Salberg, A.-B., 2020a. Dense Dilated Convolutions' Merging Network for Land Cover Classification. IEEE Trans. Geosci. Remote. Sens. 58, 6309-6320.

Liu, Q., Kampffmeyer, M., Jenssen, R., Salberg, A.B., 2020b. Dense Dilated Convolutions’ Merging Network for Land Cover Classification. IEEE Transactions on Geoscience and Remote Sensing 58, 6309-6320.

Liu, Q., Li, Z., Shuai, S., Sun, Q., 2020c. Spectral group attention networks for hyperspectral image classification with spectral separability analysis. Infrared Physics & Technology 108, 103340.

Liu, W., Liu, Z., Yu, Z., Dai, B., Lin, R., Wang, Y., Rehg, J.M., Song, L.J.I.C.C.o.C.V., Recognition, P., 2018. Decoupled Networks. 2771-2779.

Liu, X., Deng, Z., Yang, Y., 2019a. Recent progress in semantic image segmentation. Artificial Intelligence Review 52, 1089-1106.

Liu, X., Jia, S., Xu, M., Zhu, J., 2019b. Statistical Fusion-Based Transfer Learning for Hyperspectral Image Classification, IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 1108-1111.

Liu, Z., Li, X., Luo, P., Loy, C.C., Tang, X., 2015. Semantic Image Segmentation via Deep Parsing Network.

Long, J., Shelhamer, E., Darrell, T., 2015a. Fully Convolutional Networks for Semantic Segmentation.

Long, M., Cao, Y., Wang, J., Jordan, M.I., 2015b. Learning Transferable Features with Deep Adaptation Networks. arXiv:1502.02791 2.

Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik 29, 102-127.

Luo, P., Peng, Z., Ren, J., Zhang, R., 2018. Do Normalization Layers in a Deep ConvNet Really Need to Be Distinct? CoRR abs/1811.07727.

Luo, X., Shu, B., Yu, H., Wang, H., Mu, J., 2019a. A Multi-scale Network based on Attention Mechanism for Hyperspectral Image Classification, pp. 260-265.

Luo, Y., Zhu, L.-Z., Lu, B.-L., 2019b. A GAN-Based Data Augmentation Method for Multimodal Emotion Recognition, Advances in Neural Networks – ISNN 2019, pp. 141-150.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing 152, 166-177.

Ma, X., Wang, H., Geng, J., 2016. Spectral–Spatial Classification of Hyperspectral Image Based on Deep Auto-Encoder. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 4073-4085.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark, 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3226-3229.

Marchesi, M., May 2017. Megapixel Size Image Creation using Generative Adversarial Networks. arXiv:1706.00082 1.

Mayer, N., Ilg, E., Hausser, P., Fischer, P., 2015. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. arXiv:1512.02134 1.

Mei, S., Ji, J., Hou, J., Li, X., Du, Q., 2017. Learning Sensor-Specific Spatial-Spectral Features of Hyperspectral Images via Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing 55, 4520-4533.

Mnih, V., 2013. Machine Learning for Aerial Image Labeling. University of Toronto.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. Nature 518, 529-533.

Mottaghi, R., Chen, X., Liu, X., Cho, N., Lee, S., Fidler, S., Urtasun, R., Yuille, A., 2014. The Role of Context for Object Detection and Semantic Segmentation in the Wild, 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 891-898.

Mou, L., Ghamisi, P., Zhu, X.X., 2018. Unsupervised Spectral–Spatial Feature Learning via Deep Residual Conv–Deconv Network for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing 56, 391-406.

Mu, C., Guo, Z., Liu, Y., 2020. A Multi-Scale and Multi-Level Spectral-Spatial Feature Fusion Network for Hyperspectral Image Classification. Remote. Sens. 12, 125.

Nalepa, J., Antoniak, M., Myller, M., Ribalta Lorenzo, P., Marcinkiewicz, M., 2020a. Towards resource-frugal deep convolutional neural networks for hyperspectral image segmentation. Microprocessors and Microsystems 73, 102994.

Nalepa, J., Myller, M., Kawulok, M., 2019. Hyperspectral Data Augmentation.

Nalepa, J., Myller, M., Kawulok, M., 2020b. Training- and Test-Time Data Augmentation for Hyperspectral Image Segmentation. IEEE Geoscience and Remote Sensing Letters 17, 292-296.

Neuhold, G., Ollmann, T., Bul`o, S.R., Kontschieder, P., 2017. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. ICCV, 10.

Nogueira, K., Penatti, O., dos Santos, J., 2016. Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification. Pattern Recognition 61.

Park, J., Shin, C., Kim, C., 2019. PESSN: Precision Enhancement Method for Semantic Segmentation Network, 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 1-4.

Patki, N., Wedge, R., Veeramachaneni, K., 2019. The Synthetic data vault.

Penatti, O.A.B., Nogueira, K., Santos, J.A.d., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?, 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 44-51.

Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large Kernel Matters — Improve Semantic Segmentation by Global Convolutional Network, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1743-1751.

Peng, X., Sun, B., Ali, K., Saenko, K., 2015. Learning Deep Object Detectors from 3D Models, 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1278-1286.

Pepik, B., Stark, M., Gehler, P., Schiele, B., 2012. Teaching 3D geometry to deformable part models, 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3362-3369.

Pesaresi, M., Gerhardinger, A., 2011. Improved Textural Built-Up Presence Index for Automatic Recognition of Human Settlements in Arid Regions With Scattered Vegetation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 4, 16-26.

Pont-Tuset, J., Perazzi, F., Caelles, S., Arbeláez, P., Sorkine-Hornung, A., Gool, L.V., 2017. The 2017 DAVIS Challenge on Video Object Segmentation. arXiv eprint 1704.00675.

Prasad, S., Li, W., Fowler, J.E., Bruce, L.M.J.I.T.o.G., Sensing, R., 2012. Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification. 50, 3474-3486.

Qi, X., Li, K., Liu, P., Zhou, X., Sun, M., 2020. Deep Attention and Multi-Scale Networks for Accurate Remote Sensing Image Segmentation. IEEE Access 8, 146627-146639.

Radford, A., Metz, L., 2016. UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS, in: Chintala, S. (Ed.), ICLR. arXiv.

Rajpura, P.S., Bojinov, H., Hegde, R.S., 2017. Object Detection Using Deep CNNs Trained on Synthetic Images. arXiv:1706.06782 2.

Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN Features off-the-shelf: an Astounding Baseline for Recognition, CVPR. Computer Vision Foundation. .

Richter, S.R., Vineet, V., Roth, S., Koltun, V., 2016. Playing for Data: Ground Truth from Computer Games. arXiv:1608.02192 1.

Rizvi, I.A., Mohan, B.K., 2011. Object-Based Image Analysis of High-Resolution Satellite Images Using Modified Cloud Basis Function Neural Network and Probabilistic Relaxation Labeling Process. IEEE Transactions on Geoscience and Remote Sensing 49, 4815-4820.

Rodarmel, C., Shan, J.J.S., Science, L.I., 2002. Principal component analysis for hyperspectral image classification. 62, 115-122.

Romero, A., Gatta, C., Camps-Valls, G., 2016. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. IEEE Transactions on Geoscience and Remote Sensing 54, 1349-1362.

Ros, G., Ramos, S., Granados, M., Bakhtiary, A., Vazquez, D., Lopez, A.M., 2015. Vision-Based Offline-Online Perception Paradigm for Autonomous Driving, 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 231-238.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M., 2016. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3234-3243.

Rottensteiner, F., Sohn, G., Jung, J., Gerke, M., Baillard, C., Benitez, S., Breitkopf, U., 2012. The Isprs Benchmark on Urban Object Classification and 3d Building Reconstruction. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences I-3, 293-298.

Russell, B., Torralba, A., Murphy, K., Freeman, W., 2008. LabelMe: A Database and Web-Based Tool for Image Annotation. International Journal of Computer Vision 77.

S, B., 2007. CBCL streetscenes challenge framework.

Salimans, T., Kingma, D.P., 2016a. Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks, p. 901.

Salimans, T., Kingma, D.P., 2016b. Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks. arXiv:1602.07868 3.

Satkin, S., Lin, J., Hebert, M., 2012. Data-Driven Scene Understanding from 3D Models, BMVC.

Saxena, N., K. B, N., Raman, B., 2020. Semantic Segmentation of Multispectral Images using Res-Seg-net Model, 2020 IEEE 14th International Conference on Semantic Computing (ICSC), pp. 154-157.

Scharwächter, T., Enzweiler, M., Franke, U., Roth, S., 2013. Efficient Multi-cue Scene Segmentation, in: Weickert, J., Hein, M., Schiele, B. (Eds.), Pattern Recognition. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 435-445.

Seo, J., Jeon, S., Jeon, T., 2018. Domain Adaptive Generation of Aircraft on Satellite Imagery via Simulated and Unsupervised Learning.

Shen, X., Hertzmann, A., Jia, J., Paris, S., Price, B., Shechtman, E., Sachs, I., 2016. Automatic portrait segmentation for image stylization. Computer Graphics Forum 35, 93–102.

Sheng, G., Yang, W., Xu, T., Sun, H., 2012. High-resolution satellite scene classification using a sparse coding based multiple feature combination. International Journal of Remote Sensing 33, 2395-2412.

Signoroni, A., Savardi, M., Baronio, A., Benini, S., 2019a. Deep Learning Meets Hyperspectral Image Analysis: A Multidisciplinary Review. J. Imaging 5, 52.

Signoroni, A., Savardi, M., Baronio, A., Benini, S.J.J.o.I., 2019b. Deep learning meets hyperspectral image analysis: a multidisciplinary review. 5, 52.

Silberman, N., Hoiem, D., Kohli, P., Fergus, R., 2012. Indoor Segmentation and Support Inference from RGBD Images, European Conference on Computer Vision, pp. 746–760.

Simonyan, K., Zisserman, A.J.C., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. abs/1409.1556.

Sohail, A.K.a.A., Zahoora, U., Qureshi, A.S., 2020. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. Artificial Intelligence Review, 1 - 62.

Song, J., Ma, L., 2017. Deep neural network-based domain adaptation for classification of remote sensing images. Journal of Applied Remote Sensing 11.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929-1958.

Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C., 2011. The German Traffic Sign Recognition Benchmark: A multi-class classification competition, The 2011 International Joint Conference on Neural Networks, pp. 1453-1460.

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J., Wu, Z., Ding, X., 2019. Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation. arXiv.

Takikawa, T., Acuna, D., Jampani, V., Fidler, S., 2019. Gated-SCNN: Gated Shape CNNs for Semantic Segmentation.

Tao, C., Pan, H., Li, Y., Zou, Z., 2015. Unsupervised Spectral–Spatial Feature Learning With Stacked Sparse Autoencoder for Hyperspectral Imagery Classification. IEEE Geoscience and Remote Sensing Letters 12, 2438-2442.

Tao, C., Wang, H., Qi, J., Li, H., 2020. Semisupervised Variational Generative Adversarial Networks for Hyperspectral Image Classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 914-927.

Thoma, M.J.a.p.a., 2016. A survey of semantic segmentation.

Tighe, J., Lazebnik., S., 2010. Superparsing: Scalable nonparametric image parsing with superpixels. ECCV.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P., 2017. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. arXiv.

Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., Birchfield, S., 2018. Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1082-10828.

Tuia, D., Moser, G., Saux, B.L., 2017. 2016 IEEE GRSS Data Fusion Contest: Multitemporal Very High Resolution from Space [Technical Committees]. IEEE Geoscience and Remote Sensing Magazine 5, 66-70.

Tuia, D., Persello, C., Bruzzone, L., 2016a. Domain Adaptation for the Classification of Remote Sensing Data. ieee Geoscience and remote sensinG maGazine

Tuia, D., Persello, C., Bruzzone, L., 2016b. Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. IEEE Geoscience and Remote Sensing Magazine 4, 41-57.

Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T., 2014. Deep Domain Confusion: Maximizing for Domain Invariance. arXiv:1412.3474 1.

Vakalopoulou, M., Bus, N., Karantzalos, K., Paragios, N., 2017. Integrating edge/boundary priors with classification scores for building detection in very high resolution data, 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3309-3312.

Valada, A., Oliveira, G., Brox, T., Burgard, W., 2017. Deep Multispectral Semantic Scene Understanding of Forested Environments Using Multimodal Fusion.

Varshney, P.K., Arora, M.K., 2004. Advanced image processing techniques for remotely sensed hyperspectral data. Springer Science & Business Media.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is All you Need, pp. 5998-6008.

Vo, A., Truong-Hong, L., Laefer, D.F., Tiede, D., d’Oleire-Oltmanns, S., Baraldi, A., Shimoni, M., Moser, G., Tuia, D., 2016. Processing of Extremely High Resolution LiDAR and RGB Data: Outcome of the 2015 IEEE GRSS Data Fusion Contest—Part B: 3-D Contest. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 5560-5575.

Volpi, M., Ferrari, V., 2015. Semantic segmentation of urban scenes by learning local class interactions. IEEE 1-9.

Wang, C., Zhang, L., Wei, W., Zhang, Y., 2019. Hyperspectral Image Classification With Data Augmentation and Classifier Fusion. IEEE Geoscience and Remote Sensing Letters, 1-5.

Wang, M., Deng, W., 2018. Deep Visual Domain Adaptation: A Survey. arXiv:1802.03601.

Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G., 2018. Understanding Convolution for Semantic Segmentation, 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1451-1460.

Wang, S., Bai, M., Mattyus, G., Chu, H., Luo, W., Yang, B., Liang, J., Cheverie, J., Fidler, S., Urtasun, R., 2016. TorontoCity: Seeing the World with a Million Eyes. arXiv.

Windrim, L., Melkumyan, A., Murphy, R.J., Chlingaryan, A., Ramakrishnan, R., 2018. Pretraining for Hyperspectral Convolutional Neural Network Classification. IEEE Transactions on Geoscience and Remote Sensing 56, 2798-2810.

Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D., 2016. Understanding Data Augmentation for Classification: When to Warp?, 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1-6.

Wu, H., Prasad, S., 2018. Semi-Supervised Deep Learning Using Pseudo Labels for Hyperspectral Image Classification. IEEE Transactions on Image Processing 27, 1259-1270.

Wu, Z., Shen, C., van den Hengel, A., 2019. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. Pattern Recognition 90, 119-133.

Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., 2017. AID: A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification. IEEE Trans. on Geoscience and Remote Sensing 55, 3965 - 3981.

Xia, G.-S., Yang, W., Delon, J., Gousseau, Y., Sun, H., Maître, H., 2010. Structural High-resolution Satellite Image Indexing. ISPRS Journal of Photogrammetry and Remote Sensing 2, 298-303.

Xiao, J., Owens, A., Torralba, A., 2013. SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels, 2013 IEEE International Conference on Computer Vision, pp. 1625-1632.

Xiao, T., Liu, Y., Zhou, B., Jiang, Y., Sun, J., 2018. Unified Perceptual Parsing for Scene Understanding.

Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., 2017. Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. IEEE Transactions on Geoscience and Remote Sensing PP, 1-13.

Xu, Y., Du, B., Zhang, L., Cerra, D., Pato, M., Carmona, E., Prasad, S., Yokoya, N., Hansch, R., Le Saux, B., 2019. Advanced Multi-Sensor Optical Remote Sensing for Urban Land Use and Land Cover Classification: Outcome of the 2018 IEEE GRSS Data Fusion Contest. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 1709-1724.

Xu, Y., Huang, B., 2014. Spatial and temporal classification of synthetic satellite imagery: land cover mapping and accuracy validation. Geo-spatial Information Science 17, 1-7.

Yang, C., Wang, Z., 2020. An Ensemble Wasserstein Generative Adversarial Network Method for Road Extraction From High Resolution Remote Sensing Images in Rural Areas. IEEE Access 8, 174317-174324.

Yang, G., Song, X., Huang, C., Deng, Z., Shi, J., Zhou, B., 2019. DrivingStereo: A Large-Scale Dataset for Stereo Matching in Autonomous Driving Scenarios, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 899-908.

Yang, J., Zhao, Y., Chan, J.C., 2017. Learning and Transferring Deep Joint Spectral–Spatial Features for Hyperspectral Classification. IEEE Transactions on Geoscience and Remote Sensing 55, 4729-4742.

Yang, M., Kun, Y., Zhang, C., Li, Z., Yang, K., 2018. DenseASPP for Semantic Segmentation in Street Scenes.

Yang, X., Zhang, X., Ye, Y., Lau, R.Y.K., Lu, S., Li, X., Huang, X., 2020. Synergistic 2D/3D Convolutional Neural Network for Hyperspectral Image Classification. Remote. Sens. 12, 2033.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification, Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '10.

Yi, Y., Zhang, Z., Zhang, W., Zhang, C., Li, W., Zhao, T., 2019. Semantic Segmentation of Urban Buildings from VHR Remote Sensing Imagery Using a Deep Convolutional Neural Network. MDPI remote sensing 11.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? arXiv.

You, K., Long, M., Cao, Z., Wang, J., Jordan, M.I., 2019. Universal Domain Adaptation, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2715-2724.

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N., 2018. Learning a Discriminative Feature Network for Semantic Segmentation, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1857-1866.

Yu, F., Koltun, V.J.C., 2016. Multi-Scale Context Aggregation by Dilated Convolutions. abs/1511.07122.

Yuan, Y., Chen, X., Wang, J., 2019. Object-Contextual Representations for Semantic Segmentation.

Yue, J., Zhao, W., Mao, S., Liu, H., 2015. Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. Remote Sens. Lett. 6, 468–477.

Yun, K., Yu, K., Osborne, J., Eldin, S., Huyen, A., Lu, T., 2019. Improved visible to IR image transformation using synthetic data augmentation with cycle-consistent adversarial networks.

Zhan, Y., Hu, D., Wang, Y., Yu, X., 2018. Semisupervised Hyperspectral Image Classification Based on Generative Adversarial Networks. IEEE Geoscience and Remote Sensing Letters 15, 212-216.

Zhang, C., 2018. Deep Learning for Land Cover and Land Use Classification Lancaster Environment Centre. Lancaster University.

Zhang, H., Li, Y., Jiang, Y., Wang, P., Shen, Q., Shen, C., 2019. Hyperspectral Classification Based on Lightweight 3-D-CNN With Transfer Learning. IEEE Transactions on Geoscience and Remote Sensing 57, 5813-5828.

Zhang, L., Zhang, L., Du, B., 2016a. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. IEEE Geoscience and Remote Sensing Magazine 4, 22-40.

Zhang, L., Zhang, L., Du, B., 2016b. Deep learning for remote sensing data: A technical tutorial on the state of the art. IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE 4, 22-40.

Zhang, R., Candra, S.A., Vetter, K., Zakhor, A., 2015. Sensor fusion for semantic segmentation of urban scenes, 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 1850-1857.

Zhang, R., Tang, S., Zhang, Y., Li, J., Yan, S., 2017. Scale-Adaptive Convolutions for Scene Parsing, 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2050-2058.

Zhang, X., Wang, T., Yang, Y., 2020. Hyperspectral image classification based on multi-scale residual network with attention mechanism.

Zhang, Z., Zhang, X., Peng, C., Cheng, D., Sun, J., 2018. ExFuse: Enhancing Feature Fusion for Semantic Segmentation.

Zhao, B., Zhong, Y., Xia, G., Zhang, L., 2016a. Dirichlet-Derived Multiple Topic Scene Classification Model for High Spatial Resolution Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing 54, 2108-2123.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid Scene Parsing Network, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230-6239.

Zhao, H., Zhang, Y., Liu, S., Shi, J., Loy, C.C., Lin, D., Jia, J., 2018. PSANet: Point-wise Spatial Attention Network for Scene Parsing: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part IX, pp. 270-286.

Zhao, L., Tang, P., Huo, L., 2016b. Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. Journal of Applied Remote Sensing 10.

Zhen-long, 杜., Hai-yang, 沈., Guo-mei, 宋., Xiao-li, 李., 2019. Image style transfer based on improved CycleGAN. Optics and Precision Engineering 27, 1836-1844.

Zheng, Q., Chen, J., Huang, P., Hu, R., 2019a. Urban scene semantic segmentation with insufficient labeled data. China Communications 16, 212-221.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P., 2015. Conditional Random Fields as Recurrent Neural Networks.

Zheng, Z., Zhong, Y., Wang, J., 2019b. Pop-Net: Encoder-Dual Decoder for Semantic Segmentation and Single-View Height Estimation, IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 4963-4966.

Zhong, Z., Li, J., Clausi, D.A., Wong, A., 2020. Generative Adversarial Networks and Conditional Random Fields for Hyperspectral Image Classification. IEEE Trans. Cybern. 50, 3318-3329.

Zhong, Z., Li, J., Luo, Z., Chapman, M., 2018. Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. IEEE Transactions on Geoscience and Remote Sensing 56, 847-858.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A., 2017. Scene Parsing through ADE20K Dataset, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5122-5130.

Zhou, L., Zhang, C., Wu, M., 2018. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 192-1924.

Zhu, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2018. Generative Adversarial Networks for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote. Sens. 56, 5046-5063.

Zhu, M., He, Y., He, Q., 2019. A Review of Researches on Deep Learning in Remote Sensing Application. International Journal of Geosciences 10, 1-11.

Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine 5, 8-36.

Zhuo, Y.-F., Wang, Y.-L., 2018. Multilevel Features Fusion in Deep Convolutional Neural Networks, pp. 600-610.