GEOSPATIAL ARTIFICIAL INTELLIGENCE (GEOAI) METHODS FOR REMOTE SENSING APPLICATIONS

Prof. Dr. Elif SERTEL
Istanbul Technical University
Department of Geomatics Engineering
sertele@itu.edu.tr

CONTENTS

INTRODUCTION

- The emergence of RS platforms that can capture a wide range of information and an increase in earth observation data volume recently made accessible through globally accessible service delivery platforms has facilitated access to an enormous amount of high-resolution RS images and opened a new era in deep learning-based RS research.
- RS images differ greatly from natural images since RS images contain more complex patterns and exhibit a different level of complexity due to impacts of internal (detector-based) and external (environmental conditions-based) geometric distortions that occurred during the data collection.
  - Increasing number of EO satellites
  - Relatively easier access to EO data, dataset.
  - High performance computing.

INCREASE OF EO DATA




SATELLITES IN ORBIT

Satellite Database
In-depth details on the 2,666 satellites currently orbiting Earth, including their country of origin, purpose, and other operational details.
ucsusa.org

Dates: 21 December 2021- 1 September 2021- 31 March 2021

Total number of operating satellites: 4852-4550-2666

- United States: 2944-2788-1327
- Russia: 169-167-169
- China: 499-431- 363
- Other: 1240- 1164-807

•LEO: 4078-3790-1918
•MEO: 141-139-135
•Elliptical: 59-56-59
•GEO: 574-565-564

•Total number of US satellites: 2944-2788-1327

- Civil: 30-33-30
- Commercial: 2516-2359-935
- Government: 168-167-170
- Military: 230-229-192

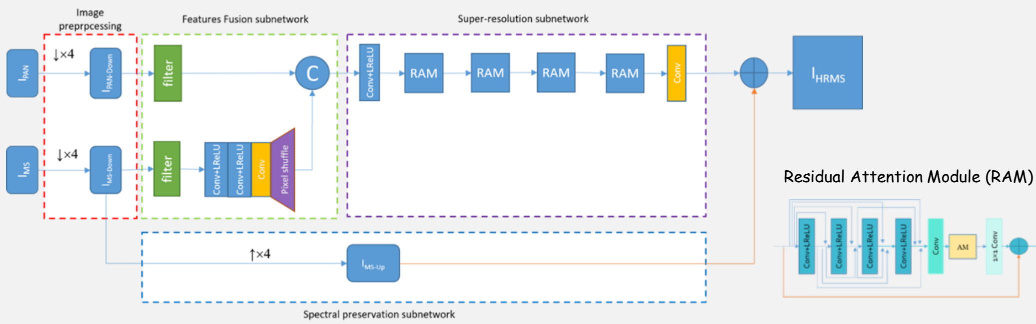
KAYNAK: https://www.ucsusa.org/resources/satellite-database

PAN-SHARPENING

- Spatial and spectral resolution of Earth observation satellites are generally not at the desired high level due to the limitations on optic and sensor technology and high costs; therefore, it is important to develop software-based algorithms to improve the spatial and spectral quality of the satellite images.
- Pan-sharpening is very important for remote sensing scene interpretation, and is also used as a pre-processing step for several image processing tasks, for example:
  - (1) feature extraction;
  - (2) segmentation;
  - (3) classification.
- The methods that we used:
  - CNN based approaches
  - Channel-spatial attention
  - GAN based approaches
  - Channel-spatial attention based GAN

CHANNEL-SPATIAL ATTENTION-BASED PAN-SHARPENING

- Aall channels are treated equally in the CNN-based networks and the abundant high-frequency information contained in the low-resolution images is not made full use of.
- The attention mechanism is proposed to address these problems and has been proven to be able to learn the deeper interdependencies among the channels



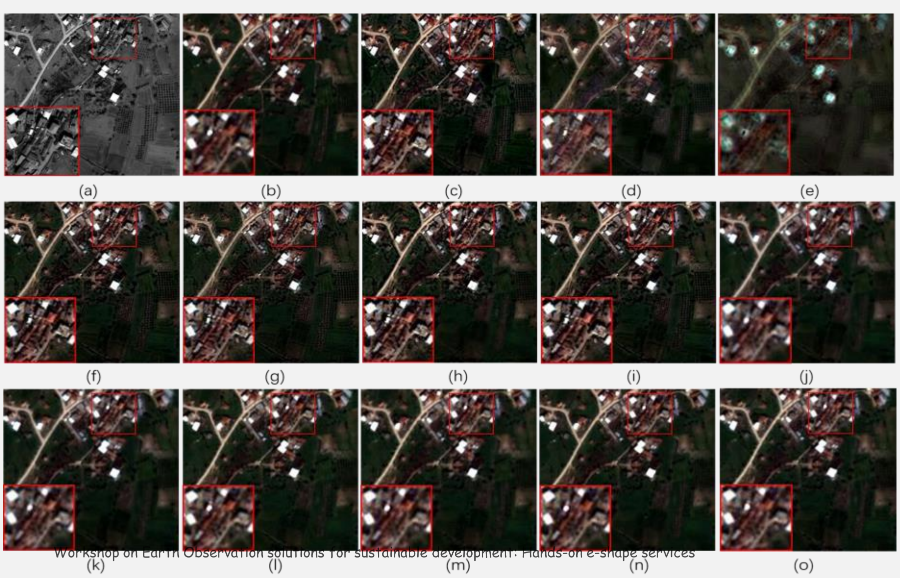
Wang, P., & Sertel, E., (2021). Channel-spatial attention-based pan-sharpening of very high-resolution satellite images. *Knowledge-Based Systems* , 1-12.

CHANNEL-SPATIAL ATTENTION-BASED PAN-SHARPENING

Train/Test	MS/PAN m	Region	Sensor	Patches		MS size	PA
Train	2/0.5	Aydin	Pleiades1A	Training: 121390 Valid: 3756		64×64	25
	2/0.5	Istanbul	Pleiades1A				
	2/0.5	Istanbul	Pleiades1A				
	2/0.5	Bursa	Pleiades1A				
	2/0.5	Mugla	Pleiades1A				
Test	2/0.5	Bursa	Pleiades1A	Reduced resolution	180	200×200	80
				Full resolution	100		
	1.6/0.4	Washington	Worldview2	Reduced resolution	80		
		Stockholm		Full resolution	65		
	2.4/0.61	Bosnia and Herzegovina	QuickBird	Reduced resolution	65		
		Amsterdam		Full resolution	50		

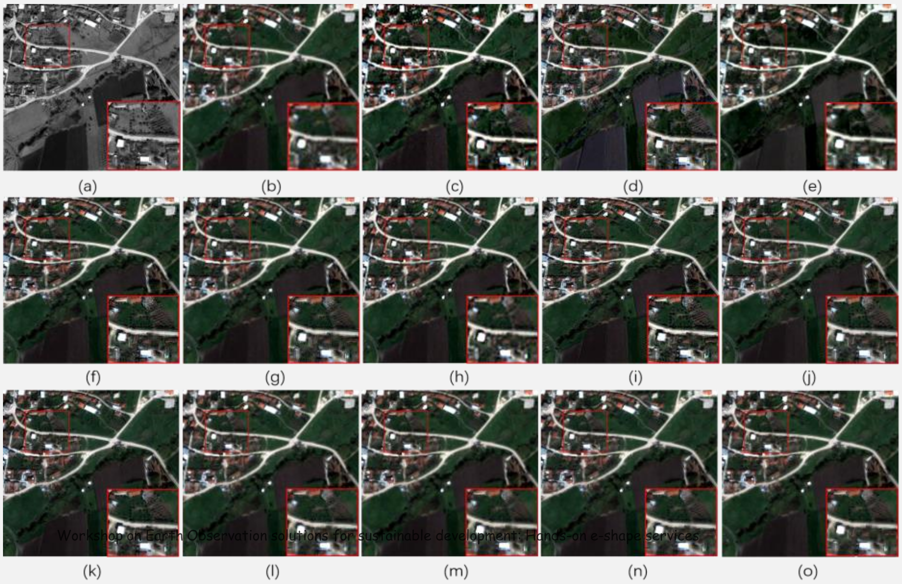
	Reduced Resolution					Full Resolution			
	QAVE	SAM	ERGAS	RMSE	CC	D_λ	D_s	QNR	GQNR
	↑	↓	↓	↓	↑	↓	↓	↑	↓
BDSD	0.931	3.327	3.613	89.866	0.915	0.027	0.083	0.893	0.137
Brovey	0.969	2.150	3.260	81.397	0.895	0.149	0.223	0.671	0.659
MTF_GLP_HPM	0.962	2.616	7.289	162.089	0.904	0.089	0.163	0.766	0.428
SFIM	0.962	2.613	7.840	158.911	0.895	0.079	0.147	0.789	0.461
SR_D	0.948	3.115	3.608	90.839	0.916	0.018	0.074	0.910	0.088
PNN	0.979	1.954	1.654	45.657	0.975	0.111	0.077	0.821	0.592
MSDCNN	0.979	1.939	1.649	45.404	0.975	0.109	0.079	0.821	0.580
PanNet	0.989	1.408	1.259	32.664	0.984	0.057	0.087	0.861	0.271
MSCARN	0.979	2.239	1.853	47.183	0.965	0.140	0.060	0.809	0.705
NCAPAN	0.984	1.643	1.427	37.866	0.980	0.041	0.100	0.862	0.203
CAPAN-CA	0.990	1.316	1.132	30.121	0.987	0.060	0.073	0.871	0.296
CAPAN-SE	0.988	1.432	1.290	33.275	0.985	0.064	**0.050**	0.890	0.363
CAPAN-CBAM	**0.991**	**1.238**	**1.066**	**28.246**	**0.988**	**0.045**	0.090	0.869	0.219

REDUCED RESOLUTION QUALITATIVE RESULTS

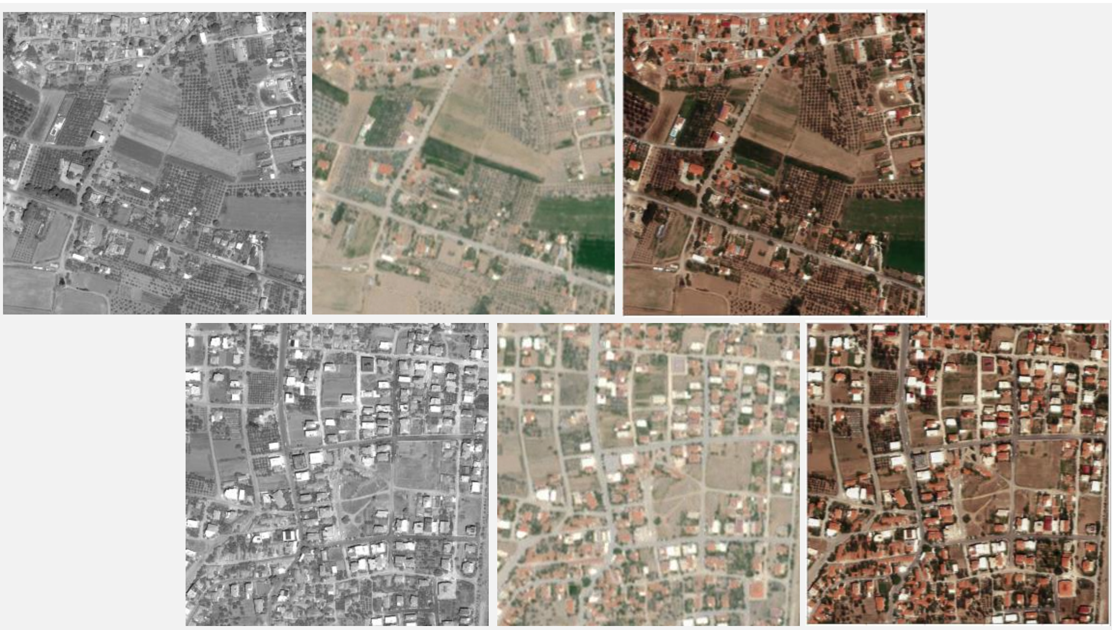


(a) Pan
(b) MS
(c) BDSD
(d) Brovey
(e) MTF_GLP_PHM
(f) SFIM
(g) SR_D
(h) PNN
(i) MSDCNN
(j) PanNet
(k) MSCARN
(l) NCAPAN
(m) CSAPAN-CA
(n) CSAPAN-SE
(o) CSAPAN-CBAM

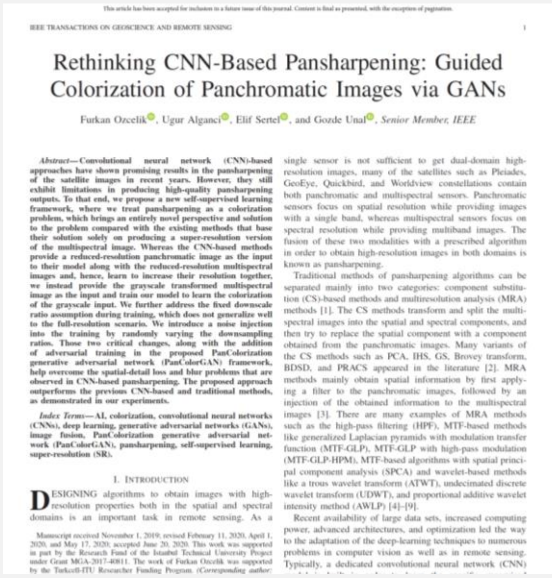
Full resolution qualitative results



(a) Pan
(b) MS
(c) BDSD
(d) Brovey
(e) MTF_GLP_PHM
(f) SFIM
(g) SR_D
(h) PNN
(i) MSDCNN
(j) PanNet
(k) MSCARN
(l) NCAPAN
(m) CSAPAN-CA
(n) CSAPAN-SE
(o) CSAPAN-CBAM



Ozcelik, F., Algancı, U., Sertel, E., & Ünal, G., (2021). Rethinking CNN-Based Pansharpening: Guided Colorization of Panchromatic Images via GANs. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* , vol.59, no.4, 3486-3501.


This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING 1

Rethinking CNN-Based Pansharpening: Guided Colorization of Panchromatic Images via GANs

Furkan Ozcelik, Ugur Alganci, Elif Sertel, and Gozde Unal, *Senior Member, IEEE*

***Abstract*—Convolutional neural network (CNN)-based approaches have shown promising results in the pansharpening of the satellite images in recent years. However, they still exhibit limitations in producing high-quality pansharpening outputs. To that end, we propose a new self-supervised learning framework, where we treat pansharpening as a colorization problem, which brings an entirely novel perspective and solution to the problem compared with the existing methods that base their solution solely on producing a super-resolution version of the multispectral image. Whereas the CNN-based methods provide a reduced-resolution panchromatic image as the input to their model along with the reduced-resolution multispectral images and, hence, learn to increase their resolution together, we instead provide the grayscale transformed multispectral image as the input and train our model to learn the colorization of the grayscale input. We further address the fixed downscale ratio assumption during training, which does not generalize well to the full-resolution scenario. We introduce a noise injection into the training by randomly varying the downsampling ratios. Those two critical changes, along with the addition of adversarial training in the proposed PanColorization generative adversarial network (PanColorGAN) framework, help overcome the spatial-detail loss and blur problems that are observed in CNN-based pansharpening. The proposed approach outperforms the previous CNN-based and traditional methods, as demonstrated in our experiments.**

***Index Terms*—AI, colorization, convolutional neural networks (CNNs), deep learning, generative adversarial networks (GANs), image fusion, PanColorization generative adversarial network (PanColorGAN), pansharpening, self-supervised learning, super-resolution (SR).**

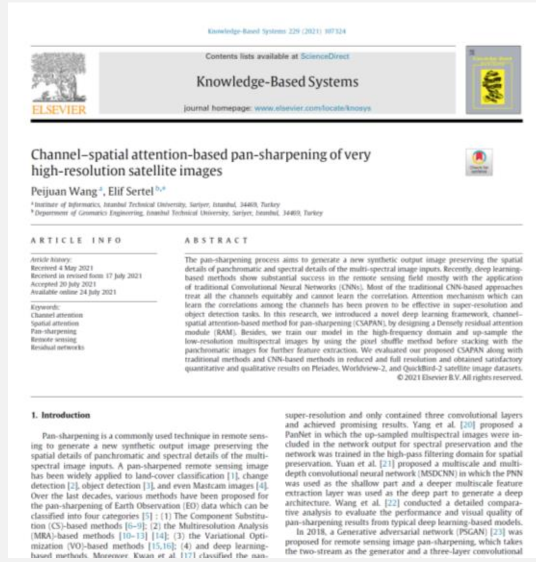
I. INTRODUCTION

DESIGNING algorithms to obtain images with high-resolution properties both in the spatial and spectral domains is an important task in remote sensing. As a single sensor is not sufficient to get dual-domain high-resolution images, many of the satellites such as Pleiades, GeoEye, Quickbird, and Worldview constellations contain both panchromatic and multispectral sensors. Panchromatic sensors focus on spatial resolution while providing images with a single band, whereas multispectral sensors focus on spectral resolution while providing multiband images. The fusion of these two modalities with a prescribed algorithm in order to obtain high-resolution images in both domains is known as pansharpening.

Traditional methods of pansharpening algorithms can be separated mainly into two categories: component substitution (CS)-based methods and multiresolution analysis (MRA) methods [1]. The CS methods transform and split the multispectral images into the spatial and spectral components, and then try to replace the spatial component with a component obtained from the panchromatic images. Many variants of the CS methods such as PCA, IHS, GS, Brovey transform, BDSD, and PRACS appeared in the literature [2]. MRA methods mainly obtain spatial information by first applying a filter to the panchromatic images, followed by an injection of the obtained information to the multispectral images [3]. There are many examples of MRA methods such as the high-pass filtering (HPF), MTF-based methods like generalized Laplacian pyramids with modulation transfer function (MTF-GLP), MTF-GLP with high-pass modulation (MTF-GLP-HPM), MTF-based algorithms with spatial principal component analysis (SPCA) and wavelet-based methods like a trous wavelet transform (ATWT), undecimated discrete wavelet transform (UDWT), and proportional additive wavelet intensity method (AWLP) [4]–[9].

Recent availability of large data sets, increased computing power, advanced architectures, and optimization led the way to the adaptation of the deep-learning techniques to numerous problems in computer vision as well as in remote sensing. Typically, a dedicated convolutional neural network (CNN)

Manuscript received November 1, 2019; revised February 11, 2020, April 1, 2020, and May 17, 2020; accepted June 20, 2020. This work was supported in part by the Research Fund of the Istanbul Technical University Project under Grant MGA-2017-40811. The work of Furkan Ozcelik was supported by the Turkcell-ITU Researcher Funding Program. (Corresponding author:


Knowledge-Based Systems 229 (2021) 107324

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

ELSEVIER

Channel–spatial attention-based pan-sharpening of very high-resolution satellite images

Peijuan Wang [a], Elif Sertel [b,*]

[a] Institute of Informatics, Istanbul Technical University, Sariyer, Istanbul, 34469, Turkey

[b] Department of Geomatics Engineering, Istanbul Technical University, Sariyer, Istanbul, 34469, Turkey

A R T I C L E I N F O

Article history:
Received 4 May 2021
Received in revised form 17 July 2021
Accepted 20 July 2021
Available online 24 July 2021

Keywords:
Channel attention
Spatial attention
Pan-sharpening
Remote sensing
Residual networks

A B S T R A C T

The pan-sharpening process aims to generate a new synthetic output image preserving the spatial details of panchromatic and spectral details of the multi-spectral image inputs. Recently, deep learning-based methods show substantial success in the remote sensing field mostly with the application of traditional Convolutional Neural Networks (CNNs). Most of the traditional CNN-based approaches treat all the channels equitably and cannot learn the correlation. Attention mechanism which can learn the correlations among the channels has been proven to be effective in super-resolution and object detection tasks. In this research, we introduced a novel deep learning framework, channel–spatial attention-based method for pan-sharpening (CSAPAN), by designing a Densely residual attention module (RAM). Besides, we train our model in the high-frequency domain and up-sample the low-resolution multispectral images by using the pixel shuffle method before stacking with the panchromatic images for further feature extraction. We evaluated our proposed CSAPAN along with traditional methods and CNN-based methods in reduced and full resolution and obtained satisfactory quantitative and qualitative results on Pleiades, Worldview-2, and QuickBird-2 satellite image datasets.
© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Pan-sharpening is a commonly used technique in remote sensing to generate a new synthetic output image preserving the spatial details of panchromatic and spectral details of the multi-spectral image inputs. A pan-sharpened remote sensing image has been widely applied to land-cover classification [1], change detection [2], object detection [3], and even Mastcam images [4]. Over the last decades, various methods have been proposed for the pan-sharpening of Earth Observation (EO) data which can be classified into four categories [5] : (1) The Component Substitution (CS)-based methods [6–9]; (2) the Multiresolution Analysis (MRA)-based methods [10–13] [14]; (3) the Variational Optimization (VO)-based methods [15,16]; (4) and deep learning-based methods. Moreover, Kwan et al. [17] classified the pan-

super-resolution and only contained three convolutional layers and achieved promising results. Yang et al. [20] proposed a PanNet in which the up-sampled multispectral images were included in the network output for spectral preservation and the network was trained in the high-pass filtering domain for spatial preservation. Yuan et al. [21] proposed a multiscale and multi-depth convolutional neural network (MSDCNN) in which the PNN was used as the shallow part and a deeper multiscale feature extraction layer was used as the deep part to generate a deep architecture. Wang et al. [22] conducted a detailed comparative analysis to evaluate the performance and visual quality of pan-sharpening results from typical deep learning-based models.

In 2018, a Generative adversarial network (PSGAN) [23] was proposed for remote sensing image pan-sharpening, which takes the two-stream as the generator and a three-layer convolutional

SUPER-RESOLUTION

- Super-resolution refers to the process of recovering high-resolution images from low-resolution images.
- Super-resolution is an important class of image processing techniques in computer vision and image processing.
- Nowadays, in the Remote Sensing field, Satellite imagery has been widely applied in agriculture, land cover classification, building extraction, prediction of disasters, and object detection.
- However, due to the limitations on the sensor technology and high costs, the imaging resolution always fails to meet the application requirement. Thus, Super-resolution is quite significant in the Remote Sensing field.

Satellite	MS/PAN (m)	region	image size(MS)	numbers of images	Data source
Pleiades	2/0.5	Aydin, Turkey	10099*10001	1	Provided by ITU-CSCRS
		Istanbul, Turkey	11523*10307	1	
		Istanbul, Turkey	10161*10543	1	
		Bursa, Turkey	10122*10695	1	
		Bursa, Turkey	10040*10136	1	
		Mugla, Turkey	9795*8840	1	
WV-2	1.6/0.4	Amsterdam	1460*2605	1	Provided by Maxar through https://earth.esa.int/eogateway
		Moen, Norway	2650*1684	1	
		Stockholm, Sweden	8391*6967	1	
		Washington	6092*6144	1	
WV-3	1.24/0.31	Tripoli, Libya	2020*1870	1	
		Rio, Brazil	3408*2794	1	
		Aksu, Turkey	4813*4980	1	Provided by Maxar
		Kestel, Turkey	4431*4920	1	
		Soganli, Turkey	3067*3987	1	
		Vegas, Nevada	162*162*8	3851	Provided by Maxar through SpaceNet2
		Paris, France	162*162*8	1149	
		Shanghai, China	163*163*8	4583	
		Khartoum, Sudan	163*163*8	1012	
		Moscow, Russia	325*325*8	1353	Provided by Maxar through SpaceNet5
		Mumbai, India	325*325*8	1016	
QuickBird-2	2.4/0.61	IJmeer, Netherlands	2938*2630	1	https://earth.esa.int/eogateway
		Nairobi, Kenya	2806*4365	1	
		Amsterdam,	2010*2615	1	
GeoEye-1	1.65/0.4	Susah, Libya	3612*1767	1	
		Donetsk Oblast, Ukraine	2192*2527	1	
		Amsterdam	3253*3528	1	
		County Waterford, Ireland	2938*2630	1	
DEIMOS	4.0/1.0	Voncauver, Canada	2928*3249	1	IEEE Data Fusion Contest 2016
		Voncauver, Canada	873*1311	1	
WV-2/WV-3			128*128	431	Pansharpening data

Details of the proposed Multi-sensor Remote Sensing Dataset (MSRSD)

RESULTS COMPARISON OF SR METHODS

Method		PSNR	SSIM	AG	LPIPS	NIQE	PI
CNN	SRCNN	35.5458	0.9961	7.5978	0.1041	5.1356	4.0781
	VDSR	35.9257	0.9965	7.6702	0.0977	4.9027	3.9546
	LGCNet	35.6082	0.9963	7.5776	0.1032	4.9646	3.9894
	PECNN	35.7425	0.9965	7.6609	0.0997	4.9660	3.9906
	RDN	36.1849	0.9968	7.7036	0.0936	4.6828	3.8375
	DDRN	36.2606	0.9968	7.6390	0.0953	4.7043	3.8582
GAN	SRGAN	34.7511	0.9949	7.6578	0.1079	4.7076	3.8617
	ESRGAN	34.2614	0.9930	7.8306	0.1112	4.5446	3.7839
	EEGAN	35.7941	0.9965	7.6951	0.0982	4.7431	3.8747
Attention-based	RCAN	36.2514	0.9968	7.6905	0.0921	4.8119	3.9115
	RSRCAN	36.1495	0.9967	7.6989	0.0934	5.0348	4.0255
	HAN	36.1936	0.9968	7.6860	0.0931	4.7927	3.8949
	SAN	36.1087	0.9967	7.6785	0.0954	4.8025	3.9019
	MHAN	36.0242	0.9967	7.6892	0.0933	4.7257	3.8587
	CARS	36.0513	0.9966	7.6167	0.0984	4.8822	3.9523
	CAFRN	35.7402	0.9965	7.6992	0.1015	4.7874	3.8918
	NLSN	36.1589	0.9967	7.6595	0.0938	4.8911	3.9493
Back-projection	DDBPN	36.1001	0.9967	7.6819	0.0952	4.7039	3.8522

(a) LR input (b) HR (c) Bic (d) SRCNN (e) VDSR (f) LGCNet (g) PECNN

(h)RDN (i) DDRN (j) SRGAN (k) ESRGAN (l) EEGAN (m) RCAN (n) RSRCAN

(o)HAN (p) SAN (q) MHAN (r) CARS (s) CAFRN (t) NLSN (u) DDBPN

LAND USE/LAND COVER CLASSIFICATION

- To categorize the given RS image with physical cover and man-made structures based on the image content.
- LCLU classification has a wide range of real-world applications such as environment monitoring, geospatial object detection, urbanization, and natural disaster analysis
  - Snapshot Ensembles (SE)
  - Stochastic Weight Averaging (SWA) and
  - Fast Geometric Ensemble (FGE)
  - InceptionResNetV3
  - SGD optimizer
  - DNNs weights were initialized

  using the He initialization technique.

Datasets	Spatial Resolution	Image size	Classes	Total Images	RS Source
NWPU-RESISC45	0.2m-30m	256x256	45	31.500	Google Earth
AID	0.5m-8m	600x600	30	10.000	Google Earth

LAND USE/LAND COVER CLASSIFICATION

- DNNE enables improvement of the performance of DNNs by ensuring the diversity of the models that are combined. Thus, enhances the generalizability of the models and produces more robust and generalizable outcomes for LCLU classification tasks.

		Overall Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
NWPU-RESISC45	Baseline	91.36	91.63	91.36	91.49
	SE	94.17	94.20	94.17	94.19
	SWA	**94.51**	**94.51**	**94.47**	**94.49**
	FGE	93.33	93.41	93.33	93.37
AID	Baseline	94.46	94.51	94.56	94.53
	SE	95.50	95.39	95.24	95.31
	SWA	**97.42**	**97.34**	**97.28**	**97.30**
	FGE	94.64	94.46	94.40	94.43

INDEX	CLASSES	BASE-PA	SWA-PA	DIFF-PA	BASE-UA	SWA-UA	DIFF-UA
1	airplane	97,12	98,73	1,60	96,43	96,96	0,54
2	airport	91,41	91,25	-0,16	91,25	95,00	3,75
3	baseball_diamond	97,43	97,30	-0,13	94,82	96,43	1,61
4	basketball_court	96,64	97,29	0,65	92,50	96,07	3,57
5	beach	89,37	95,31	5,94	99,11	98,04	-1,07
6	bridge	95,65	97,58	1,93	94,29	93,57	-0,71
7	chaparral	93,60	97,55	3,95	99,29	99,46	0,18
8	church	80,97	86,67	5,70	77,50	76,61	-0,89
9	circular_farmland	98,02	98,58	0,55	97,32	98,93	1,61
10	cloud	98,53	98,75	0,21	95,89	98,39	2,50
11	commercial_area	74,22	88,10	13,89	93,04	91,25	-1,79
12	dense_residential	84,94	87,93	2,99	83,57	92,32	8,75
13	desert	83,28	94,27	10,99	92,50	93,93	1,43
14	forest	93,47	92,64	-0,84	94,64	96,61	1,96
15	freeway	93,12	95,27	2,15	89,46	90,00	0,54
16	golf_course	95,99	96,35	0,37	98,21	99,11	0,89
17	ground_track_field	96,44	97,46	1,02	96,79	96,07	-0,71
18	harbor	96,63	97,72	1,08	97,32	99,29	1,96
19	industrial_area	81,85	93,35	11,50	86,96	90,18	3,21
20	intersection	97,00	98,47	1,48	92,32	92,14	-0,18
21	island	98,66	97,13	-1,53	91,79	96,61	4,82
22	lake	90,94	92,40	1,46	91,43	93,39	1,96
23	meadow	90,80	98,28	7,48	93,39	91,79	-1,61
24	medium_residential	80,57	89,48	8,91	80,71	86,61	5,89
25	mobile_home_park	83,78	94,58	10,81	95,89	96,61	0,71
26	mountain	95,15	92,51	-2,63	80,54	92,68	12,14
27	overpass	95,85	93,28	-2,57	90,71	96,61	5,89
28	palace	83,26	78,91	-4,35	63,93	80,18	16,25
29	parking_lot	95,18	95,01	-0,17	95,18	98,57	3,39
30	railway	81,98	87,19	5,21	91,79	94,82	3,04
31	railway_station	90,06	88,30	-1,76	82,50	91,61	9,11
32	rectangular_farmland	95,29	93,38	-1,91	79,46	93,21	13,75
33	river	94,97	94,58	-0,39	84,29	90,36	6,07
34	roundabout	98,00	93,89	-4,11	96,25	98,75	2,50
35	runway	95,50	95,67	0,17	90,89	90,71	-0,18
36	sea_ice	99,28	99,27	0,01	97,86	98,39	0,54
37	ship	93,29	96,94	3,66	94,29	96,25	1,96
38	snowberg	95,54	97,70	2,16	95,54	98,57	3,04
39	sparse_residential	89,61	92,09	2,48	93,93	93,57	-0,36
40	stadium	96,42	94,09	-2,33	91,43	96,61	5,18
41	storage_tank	96,17	96,49	0,32	94,11	98,21	4,11
42	tennis_court	93,41	93,22	-0,19	96,25	98,21	1,96
43	terrace	79,94	92,38	12,43	97,50	93,04	-4,46
44	thermal_power_station	88,02	96,98	8,95	90,54	91,61	1,07
45	wetland	86,32	93,79	7,47	87,86	88,93	1,07

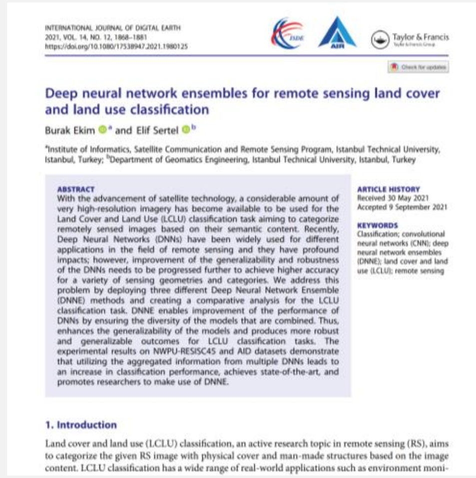
SWA is better than Baseline for Producer Accuracy	30
Baseline is better than SWA for Producer Accuracy	15
SWA is better than Baseline for User Accuracy	35
Baseline is better than SWA for User Accuracy	10

INDEX	CLASSES	BASE-P	SWA-P	DIFF-P	BASE-R	SWA-R	DIFF-R
1	airplane	0,971	0,994	0,023	0,964	0,966	0,002
2	airport	0,914	0,898	-0,017	0,913	0,970	0,057
3	baseball_diamond	0,974	0,982	0,008	0,948	0,966	0,018
4	basketball_court	0,966	0,969	0,003	0,925	0,959	0,034
5	beach	0,894	0,939	0,046	0,991	0,966	-0,025
6	bridge	0,957	0,974	0,018	0,943	0,952	0,009
7	chaparral	0,936	0,949	0,013	0,993	0,998	0,005
8	church	0,810	0,861	0,052	0,775	0,743	-0,032
9	circular_farmland	0,980	0,974	-0,007	0,973	0,991	0,018
10	cloud	0,985	0,987	0,002	0,959	0,963	0,004
11	commercial_area	0,742	0,844	0,101	0,930	0,905	-0,025
12	dense_residential	0,849	0,891	0,042	0,836	0,891	0,055
13	desert	0,833	0,949	0,117	0,925	0,870	-0,055
14	forest	0,935	0,938	0,003	0,946	0,943	-0,004
15	freeway	0,931	0,933	0,002	0,895	0,898	0,004
16	golf_course	0,960	0,953	-0,006	0,982	0,986	0,004
17	ground_track_field	0,964	0,948	-0,016	0,968	0,980	0,013
18	harbor	0,966	0,970	0,004	0,973	0,986	0,013
19	industrial_area	0,818	0,883	0,064	0,870	0,875	0,005
20	intersection	0,970	0,986	0,016	0,923	0,905	-0,018
21	island	0,987	0,942	-0,044	0,918	0,964	0,046
22	lake	0,909	0,889	-0,021	0,914	0,941	0,027
23	meadow	0,908	0,991	0,083	0,934	0,816	-0,118
24	medium_residential	0,806	0,845	0,039	0,807	0,864	0,057
25	mobile_home_park	0,838	0,949	0,111	0,959	0,959	0,000
26	mountain	0,951	0,913	-0,039	0,805	0,914	0,109
27	overpass	0,958	0,936	-0,022	0,907	0,970	0,063
28	palace	0,833	0,780	-0,052	0,639	0,780	0,141
29	parking_lot	0,952	0,925	-0,026	0,952	0,975	0,023
30	railway	0,820	0,932	0,112	0,918	0,854	-0,064
31	railway_station	0,901	0,840	-0,060	0,825	0,911	0,086
32	rectangular_farmland	0,953	0,861	-0,092	0,795	0,914	0,120
33	river	0,950	0,913	-0,037	0,843	0,934	0,091
34	roundabout	0,980	0,946	-0,034	0,963	0,993	0,030
35	runway	0,955	0,977	0,022	0,909	0,895	-0,014
36	sea_ice	0,993	0,995	0,002	0,979	0,979	0,000
37	ship	0,933	0,963	0,030	0,943	0,970	0,027
38	snowberg	0,955	0,942	-0,013	0,955	0,991	0,036
39	sparse_residential	0,896	0,870	-0,026	0,939	0,946	0,007
40	stadium	0,964	0,947	-0,017	0,914	0,963	0,048
41	storage_tank	0,962	0,958	-0,003	0,941	0,982	0,041
42	tennis_court	0,934	0,946	0,012	0,963	0,968	0,005
43	terrace	0,799	0,898	0,098	0,975	0,923	-0,052
44	thermal_power_station	0,880	0,949	0,069	0,905	0,929	0,023
45	wetland	0,863	0,959	0,096	0,879	0,795	-0,084

SWA is better than Baseline for Precision Score	27
Baseline is better than SWA for Precision Score	18
SWA is better than Baseline for Recall Score	32
Baseline is better than SWA for Recall Score	11
Equal Scores for Recall	2

DEEP NEURAL NETWORK ENSEMBLE RESULTS

Figure 2. True classifications: (a) Airport, (b) Bridge, and (c) Harbor. Misclassifications: (Ground Truth -> Prediction). (d) Church -> Palace, (e) Palace -> Church, and (f) Forest -> Wetland for the NWPU-RESISC45 dataset.



INTERNATIONAL JOURNAL OF DIGITAL EARTH
2021, VOL. 14, NO. 12, 1868–1881
https://doi.org/10.1080/17538947.2021.1980125

Deep neural network ensembles for remote sensing land cover and land use classification

Burak Ekim [a] and Elif Sertel [b]

[a]Institute of Informatics, Satellite Communication and Remote Sensing Program, Istanbul Technical University, Istanbul, Turkey; [b]Department of Geomatics Engineering, Istanbul Technical University, Istanbul, Turkey

ABSTRACT
With the advancement of satellite technology, a considerable amount of very high-resolution imagery has become available to be used for the Land Cover and Land Use (LCLU) classification task aiming to categorize remotely sensed images based on their semantic content. Recently, Deep Neural Networks (DNNs) have been widely used for different applications in the field of remote sensing and they have profound impacts; however, improvement of the generalizability and robustness of the DNNs needs to be progressed further to achieve higher accuracy for a variety of sensing geometries and categories. We address this problem by deploying three different Deep Neural Network Ensemble (DNNE) methods and creating a comparative analysis for the LCLU classification task. DNNE enables improvement of the performance of DNNs by ensuring the diversity of the models that are combined. Thus, enhances the generalizability of the models and produces more robust and generalizable outcomes for LCLU classification tasks. The experimental results on NWPU-RESISC45 and AID datasets demonstrate that utilizing the aggregated information from multiple DNNs leads to an increase in classification performance, achieves state-of-the-art, and promotes researchers to make use of DNNE.

ARTICLE HISTORY
Received 30 May 2021
Accepted 9 September 2021

KEYWORDS
Classification; convolutional neural networks (CNN); deep neural network ensembles (DNNE); land cover and land use (LCLU); remote sensing

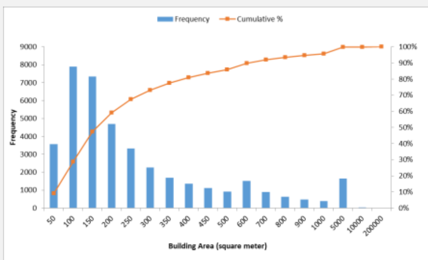
1. Introduction

Land cover and land use (LCLU) classification, an active research topic in remote sensing (RS), aims to categorize the given RS image with physical cover and man-made structures based on the image content. LCLU classification has a wide range of real-world applications such as environment moni-

BUILDING DETECTION-PROBLEMS

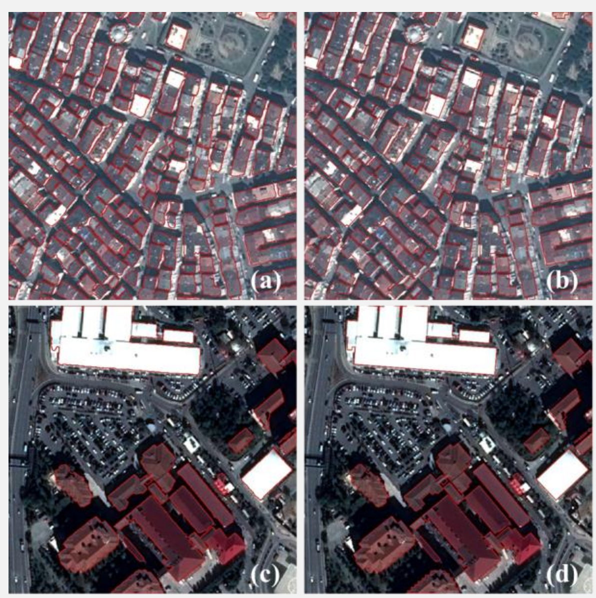
A NEW BUILDING DATASET

- 150 Pléiades image tiles of 1500 x 1500 pixels
- covering an area of 85 km^2 area of Istanbul city
- Aproximately 40,000 buildings were labelled representing different building structures and spatial distribution.

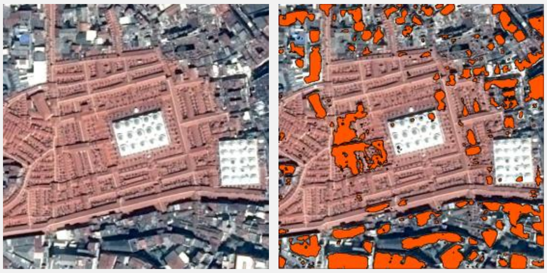


Bakirman, T., Komurcu, I. & Sertel, E., (2022) Comparative analysis of deep learning-based building extraction methods with the new VHR Istanbul dataset, Experts Systems with Applications, vol. 202, 117346

- More than 60 experiments were conducted by applying state-of-the-art architectures such as U-Net, Unet++, DeepLabv3+, FPN and PSPNet with different pre-trained encoders and hyperparameters.
- Unet++ architecture using SE-ResNeXt101 encoder pre-trained with ImageNet provides the best results with 93.8% IoU on Istanbul dataset.



CHALLENGING CONDITIONS-BAD RESULTS



CROSS EVALUATION

- Inria dataset (Maggiori et al. 2017) consist of orthorectified color imagery (RGB) with a spatial and spectral resolution of 0.3 m and 8-bit, respectively.
- The Massachusetts dataset (Mnih 2013) consists of 151 aerial RGB images with 1500 x 1500 pixels covering approximately 340 km2 from the Boston area. The images have 1 m spatial and 8-bit spectral resolution.

Train / Test	Istanbul	Inria	Massachusetts
Istanbul	**0.9380**	0.8446	0.8300
Inria	0.6121	**0.7539**	0.5338
Massachusetts	0.8473	0.8494	**0.9253**

POST-PROCESSING ON VECTOR DATA

Expert Systems With Applications 202 (2022) 117346

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Comparative analysis of deep learning based building extraction methods with the new VHR Istanbul dataset

Tolga Bakirman [a,*,1], Irem Komurcu [b,2], Elif Sertel [c,3]

[a] *Geomatics Engineering, Yildiz Technical University, Esenler 34220, Istanbul, Turkey*
[b] *Deloitte Touche Tohmatsu Limited Turkey, Istanbul, Turkey*
[c] *Geomatics Engineering, Istanbul Technical University, Istanbul, Turkey*

ARTICLE INFO

Keywords:
Building extraction
Deep learning
Pléiades
Urban

ABSTRACT

Automatic building segmentation from satellite images is an important task for various applications such as urban mapping, disaster management and regional planning. With the broader availability of very high-resolution satellite images, deep learning-based techniques have been broadly used for remote sensing image-related tasks. In this study, we generated a new building dataset, the Istanbul dataset, for the building segmentation task. 150 Pléiades image tiles of 1500 × 1500 pixels covering an area of 85 km^2 area of Istanbul city were used and approximately 40,000 buildings were labelled, representing different building structures and spatial distribution. We extensively investigated the ideal architecture, encoder and hyperparameter settings for building segmentation tasks using the new Istanbul dataset. More than 60 experiments were conducted by applying state-of-the-art architectures such as U-Net, Unet++, DeepLabv3+, FPN and PSPNet with different pre-trained encoders and hyperparameters. Our experiments showed that Unet++ architecture using SE-ResNeXt101 encoder pre-trained with ImageNet provides the best results with 93.8% IoU on the Istanbul dataset. In order to prove our solution's generalizability, the ideal network has also been trained separately on Inria and Massachusetts building segmentation datasets. The networks have produced IoU values of 75.39% and 92.53% on the Inria and Massachusetts datasets, respectively. The results indicate that our ideal network solution settings outperform other methods in terms of building segmentation even without any specific architectural modification. The weights files and inference notebook is available on: https://github.com/TolgaBkm/Istanbul_Dataset.

LU/LC MAPPING WITH MULTI-MODAL REMOTE SENSING DATA

- Each sensor captures unique information regarding the physical attribution of an earth's surface (such as geometric structure, orientation, spectral information, and roughness)
- It is chalenging to capture the various LU/LC classes using single modality data.
- Recent advances in EO and sensor technology have led to the increased accessibility of optical and SAR images at different resolutions.
- It is vital to make use of multiple sensors (multimodal) rather than a single sensor (unimodal) to get a better understanding of the area of interest.

DATASET: OPTICAL

Specifications

- Pansharpaned Multispectral product type
- 4 bands (RGBNIR)
- MS 457-874 nm and 2.4 m GSD
- PAN 250-745 with 70 cm GSD
- 1.5m spatial resolution
- bit depth: 16 (uint16)
- Acquisition Date: 13 May 2019

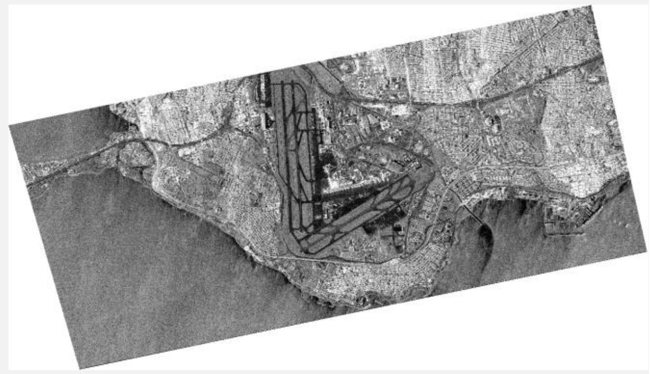
SPOT-7

DATASET: SAR

Specifications

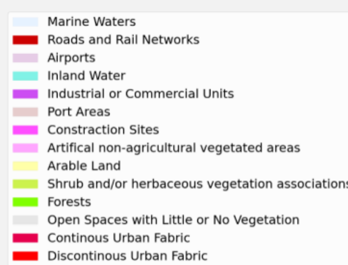
- HH polarization
- X-band frequency
- High Resolution Spotlight Mode (HS) beam mode
- Single look complex (SSC) product type.
- Spatial resolution of 1.1m in azimuth and 0.6m in range.
- 16 bit cosar (complex SAR) data type.
- Acquisition Date: 6 Jun 2019

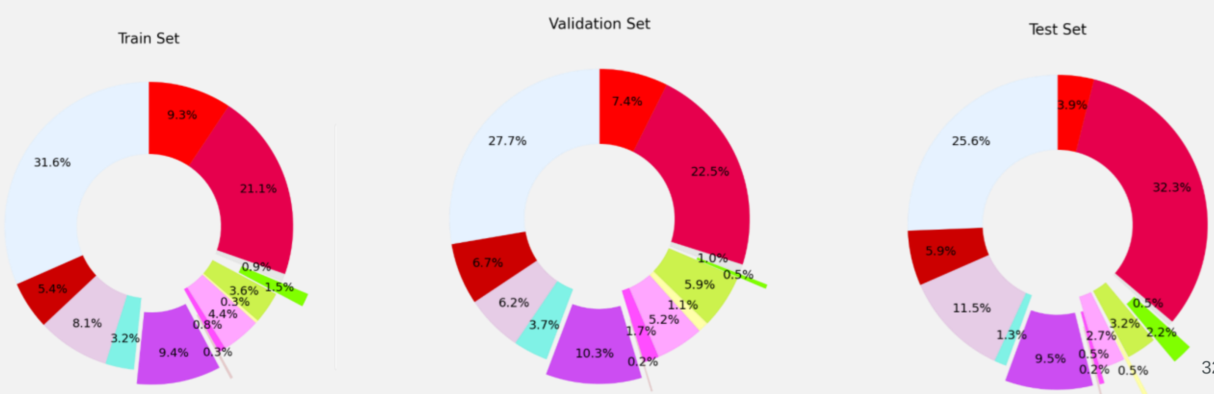
TerraSAR-X



31

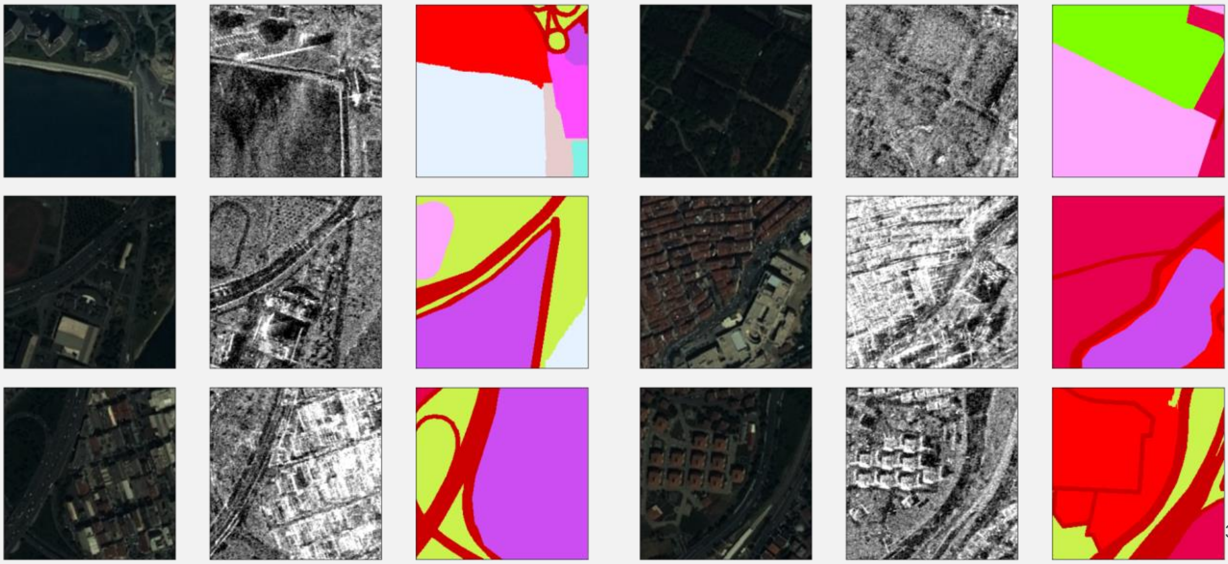
DATASET: STATISTICS





32

DATASET: SAMPLE PATCHES



Comparison of backbone architectures.

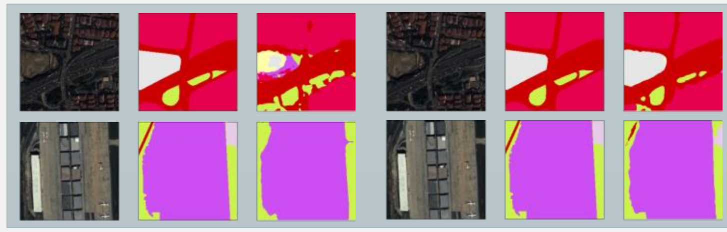
Backbone	Parameters (Millions)	IOU Score	F1 Score	Precision	Recall
ResNet 50	23	91.69	95.58	95.59	95.58
ResNeXt 50	**22**	**92.03**	**95.79**	**95.81**	**95.77**
EfficientNet B0	4	90.21	94.75	94.77	94.73
DPN68	11	91.60	95.54	95.56	95.52
MobileNet v2	2	89.39	94.30	94.33	94.28

Comparison of segmentation architectures.

Architecture	IOU	F1 Score	Precision	Recall
U-Net	79.25	87.92	87.93	87.91
U-Net ++	86.88	92.75	92.76	92.74
DeepLab v3+	**91.69**	**95.58**	**95.59**	**95.58**
MA-Net	75.21	85.06	85.06	85.01
PSPNet	88.63	93.83	93.88	93.78

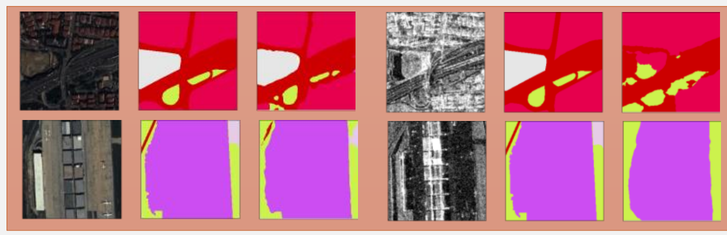
Comparison of 3-band and 4-band input

Input	IOU	F1 Score	Precision	Recall
3-band optical (RGB)	84.57	91.39	91.51	91.28
4-band optical (RGBNIR)	90.65	95.02	95.05	95



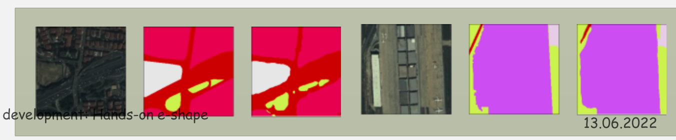
Comparison of optical and SAR data

Input	IOU	F1 Score	Precision	Recall
Optical	90.65	95.02	95.05	95
SAR	84.70	91.39	91.48	91.30



Multi-modal Learning

Fusion Method	IOU	F1 Score	Precision	Recall
Early Fusion	92.37	95.95	95.97	95.92
Middle Fusion	70.93	82.27	83.04	81.46
Late Fusion	86.73	92.79	88.65	97.43

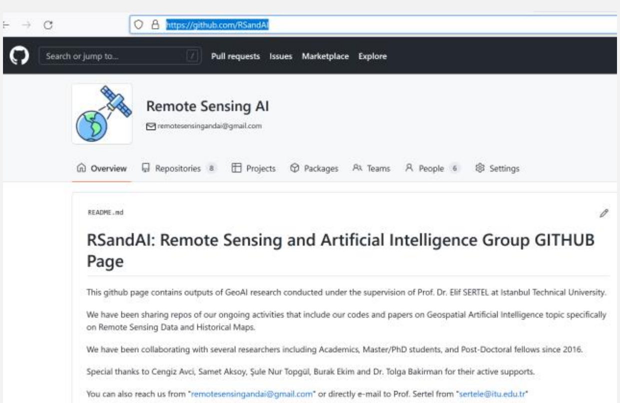


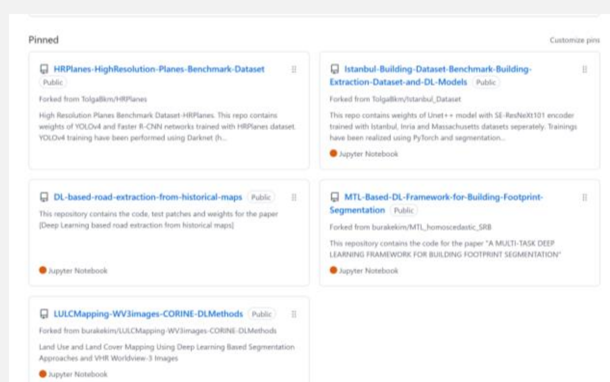
CONCLUSIONS AND SUGGESTIONS

- PAN-SHARPENING and SUPER-RESOLUTION
  - Lack of real reference data
  - Contradictory results among different spectral and spatial metrics
  - Visual control is necessary
  - Inclusion of different sensors
  - Sensor specific weights
  - Test data set should be representative of different LC/LU classes, objects of different sizes and some distict color features
  - UAV images obtained from different altitudes could be used to generate a benchmark dataset

CONCLUSIONS AND SUGGESTIONS

- LABELLING PROBLEMS
  - Accurate, highly representative of variou acquisition and landscape conditions
  - Be careful about noisy labels
- LU/LC Classification
  - Increase in number and variety of the classes cause challanges
  - Ensembles provide better results
  - Multi-label data sets..
- OBJECT DETECTION
  - Iterative creation of labelled data
  - Multi-task learning
  - Post-processing
- SEGMENTATION
  - Creating densely annotated GT is hard: Making use of semi-supervised, unsupervised, zero-shot, few-shot learning techniques could be beneficial.
  - Class unbalance.
  - It is hard to acquire EO-data: Generative models to synthesize optical/SAR/GT information might be a useful tool.
  - Multi-modal data could improve LU/LC segmentation





THANKS FOR YOUR ATTENTION...

sertele@itu.edu.tr
https://github.com/RSandAI