Firstly, I would like to thank you for this assignment. Due to the demands of my work, I had to focus on this in the last few days, but I truly enjoyed the learning process. I initially embarked on my deep learning journey with PyTorch in Udacity's Deep Learning Nanodegree program two years ago. However, since then, I have been primarily acquainted with Keras. This assignment provided me with the opportunity to revisit PyTorch and delve into new concepts such as feature indexing, Grad-CAM, and CBIR overall.

"In the assignment, initially, I inspected datasets and based on their sizes and classes, I decided to use a pre-trained model on the RSISC dataset. Considering the provided data, I opted for 4 classes and excluded the 'neighborhood' class since it had very few images, leading to an imbalance in data distribution. Given that data augmentation alone wouldn't suffice to address this imbalance issue, overcoming the problem requires further collection or injection of data from other sets.

I assumed that it is permissible to use the given datasets; otherwise, I would merge images for the selected classes from all datasets.

Since the classes I selected from the given partial datasets are strongly related to objects (unlike very general surface types), I would like to investigate if I could involve YOLO in the project.

During the fine-tuning process, I encountered a high bias problem, and I attempted to address this issue by collecting more data if possible.

I used KDTree for feature indexing as suggested and employed the Minkowski metric for distances.

In the final part, it can be observed that Grad-CAM provides attention to the objects for classification.

OVERVIEW OF DATASETS FOR DECIDING PRE-TRAINED MODEL

AID

The dataset is made up of the following 30 aerial scene types: *airport, bare land, baseball field, beach, bridge, center, church, commercial, dense residential, desert, farmland, forest, industrial, meadow, medium residential, mountain, park, parking, playground, pond, port, railway station, resort, river, school, sparse residential, square, stadium, storage tanks and viaduct*. All the images are labelled by the specialists in the field of remote sensing image interpretation, and some samples of each class are shown in Fig.1. In all, the AID dataset has a number of 10000 images within 30 classes.

BCS two classes: coffee and non-coffee

CLRS dataset consists of 15,000 remote sensing images divided into 25 scene classes, namely, *airport, bare-land, beach, bridge, commercial, desert, farmland, forest, golf-course, highway, industrial, meadow, mountain, overpass, park, parking, playground, port, railway, railway-station, residential, river, runway, stadium, and storage-tank*

BIGEARTH

Table 2: The considered Level-3 CLC classes and the number of images associated with each land-cover class in the BigEarthNet.

Land-Cover Classes	Number of Images
Mixed forest	217,119
Coniferous forest	211,703
Non-irrigated arable land	196,695
Transitional woodland/shrub	173,506
Broad-leaved forest	150,944
Land principally occupied by agriculture, with significant areas of natural vegetation	147,095
Complex cultivation patterns	107,786
Pastures	103,554
Water bodies	83,811
Sea and ocean	81,612
Discontinuous urban fabric	69,872
Agro-forestry areas	30,674
Peatbogs	23,207
Permanently irrigated land	13589
Industrial or commercial units	12895
Natural grassland	12,835
Olive groves	12,538
Sclerophyllous vegetation	11,241
Continuous urban fabric	10,784
Water courses	10,572
Vineyards	9,567
Annual crops associated with permanent crops	7,022
Inland marshes	6,236
Moors and heathland	5,890
Sport and leisure facilities	5,353
Fruit trees and berry plantations	4,754
Mineral extraction sites	4,618
Rice fields	3,793
Road and rail networks and associated land	3,384
Bare rock	3,277
Green urban areas	1,786
Beaches, dunes, sands	1,578
Sparsely vegetated areas	1,563
Salt marshes	1,562
Coastal lagoons	1,498
Construction sites	1,174
Estuaries	1,086
Intertidal flats	1,003
Airports	979
Dump sites	959
Port areas	509
Salines	424
Burnt areas	328

DFC 15

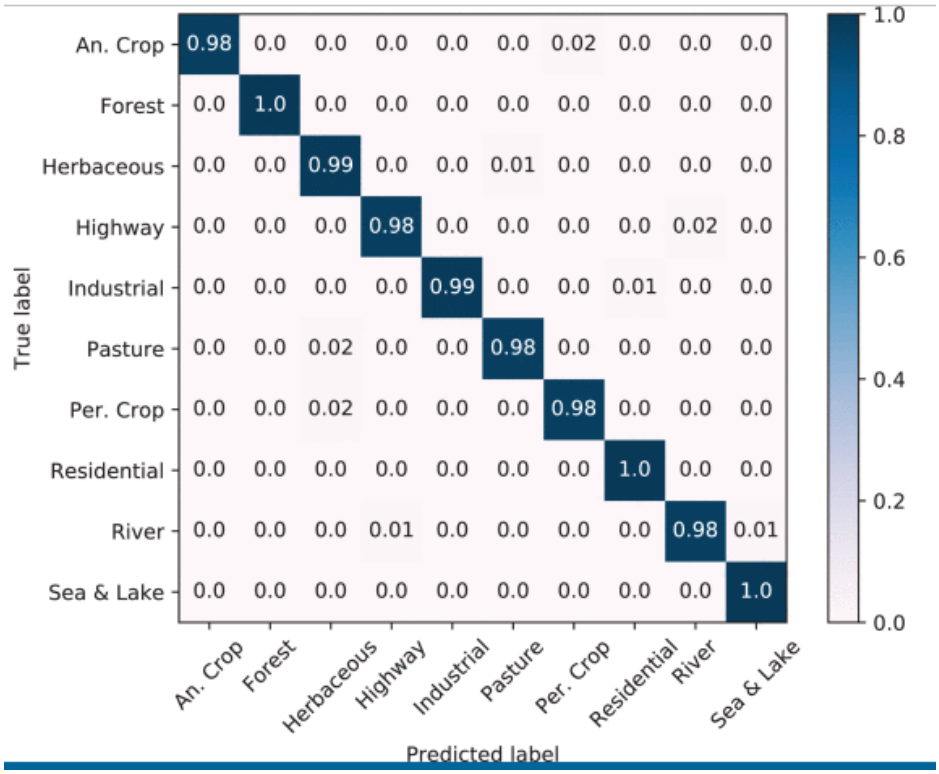
Table 1: The Number of Images in Each Object Class

Class No.	Class Name	Total	Training	Test
1	airplane	100	80	20
2	bare soil	718	577	141
3	building	691	555	136
4	car	886	722	164
5	chaparral	115	82	33
6	court	105	84	21
7	dock	100	80	20
8	field	104	79	25
9	grass	975	804	171
10	mobile home	102	82	20
11	pavement	1300	1047	253
12	sand	294	218	76
13	sea	100	80	20
14	ship	102	80	22
15	tank	100	80	20
16	tree	1009	801	208
17	water	203	161	42
-	All	2100	1680	420

Table 2: The Number of Images in Each Object Class

Class No.	Class Name	Total	Training	Test
1	impervious	3133	2532	602
2	water	998	759	239
3	clutter	1891	1801	90
4	vegetation	1086	522	562
5	building	1001	672	330
6	tree	258	35	223
7	boat	270	239	31
8	car	705	478	277
-	All	3342	2674	668

Eurosat



MLRSNet is composed of 109,161 labeled RGB images from all around the world annotated into 46 broad categories: airplane, airport, bareland, baseball diamond, basketball court, beach, bridge, chaparral, cloud, commercial area, dense residential area, desert, eroded farmland, farmland, forest, freeway, golf course, ground track field. harbor&port, industrial area, intersection, island, lake, meadow, mobile home park, mountain, overpass, park, parking lot, parkway, railway, railway station, river, roundabout, shipping yard, snowberg, sparse residential area, stadium, storage tank, swimming pool, tennis court, terrace, transmission tower, vegetable greenhouse, wetland, and wind turbine.

OPTIMAL

In this data set, 31 classes are constructed and each class is formed by 60 images with the size of
256×256 pixels, so it has a total of 1860 images. In addition, the classes of our data set includes airplane, airport, baseball field, basketball court, beach, bridge, bushes, church, round farmland, business district, dense houses, desert, forest, freeway, golf field, playground, harbor, factory, crossroads, island, lake, meadow, medium houses, mobile house area, mountain, overpass, parking lot, railway, square farmland, roundabout, and runway.

PatternNet contains 38 classes: airplane, baseball field, basketball court, beach, bridge, cemetery, chaparral, Christmas tree farm, closed road, coastal mansion, crosswalk, dense residential, ferry terminal, football field, forest, freeway, golf course, harbor, intersection, mobile home park, nursing home, oil gas field, oil well, overpass, parking lot, parking space, railway, river, runway, runway marking, shipping yard, solar panel, sparse residential, storage tank, swimming pool, tennis court, transformer station and wastewater treatment plant. There are a total of 800 images of size 256 × 256 pixels in each class. (30400)

RESISC45 dataset (https://1drv.ms/u/s!AmgKYzARBl5ca3HNaHllzp_IXjs). The NWPU-RESISC45 dataset [54] is currently the largest publicly available benchmark dataset for remote sensing scene classification. It was also recently released. It is constructed by first investigating all scene classes of the existing datasets and then selecting a list of 45 representative scene classes: airplane, airport, baseball diamond, basketball court, beach, bridge, chaparral, church, circular farmland, cloud, commercial area, dense residential, desert, forest, freeway, golf course, ground track field, harbor, industrial area, intersection, island, lake, meadow, medium residential, mobile home park, mountain, overpass, palace, parking lot, railway, railway station, rectangular farmland, river, roundabout, runway, sea ice, ship, snow berg, sparse residential,

stadium, storage tank, tennis court, terrace, thermal power station, and wetland. Each class has 700 images (31500)of size 256 × 256 pixels and the spatial resolution of the images in each class varies from about 0.2 to 30m

RSD 46

2326 images downloaded from Google Earth and Tianditu [49]. We labeled the objects in these images with four categories: oil tank, aircraft, overpass, and playground

TABLE III

STATISTICS OF THE TRAINING AND VALIDATION DATA SETS

Class	Image Number	Original Object Number	Augment Object Number	Total
Oil tank	165	1586	140498	142084
Aircraft	446	4993	221696	226689
Overpass	176	180	13636	13816
Playground	189	191	27741	27932
Background	2326	668984	0	668984

RSI-CB

This benchmark has two sub-datasets with 256 × 256 and 128 × 128 sizes because different DCNNs require different image sizes. The former contains 6 categories with 35 subclasses of more than 24,000 images. The latter contains 6 categories with 45 subclasses of more than 36,000 images. The six categories are agricultural land, construction land and facilities, transportation and facilities, water and water conservancy facilities, woodland, and other lands, and each has several subclasses.

RSSCN7 dataset (https://www.dropbox.com/s/j80iv1a0mvhonsa/RSSCN7.zip?dl=0). The RSSCN7 dataset [116] is sampled on four different scale levels from Google Earth imagery and consists of 7 classes: grassland, forest, farmland, parking lot, residential region, industrial region, river, and lake. There are 400 images in each class and each image has size of 400 × 400 pixels. (2800)

SAT-6 consists of a total of 405,000 image patches each of size 28×28 and covering 6 landcover classes - barren land, trees, grassland, roads, buildings and water bodies.

SIRI-WHU dataset comprises of 2400 images organized into 12 categories:: agriculture, commercial, harbor, idle land, industrial, meadow, overpass, park, pond, residential, river, and water. Each class contains 200 images with the size of 200 × 200 pixels

UC Merced dataset - (http://vision.ucmerced.edu/datasets/landuse.html). The UC Merced dataset (UCMD) [7] is a land use/land cover dataset which contains 100 images of the following 21 classes (2100): agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks and tennis courts. Each image measures 256 × 256 pixels

WHU-RS19 dataset (http://dsp.whu.edu.cn/cn/staff/yw/HRSscene.html). The WHURS19 remote sensing dataset (RSD) [115] is manually collected from Google Earth Imagery and labeled into 19 classes: airport, beach, bridge, commercial area, desert, farmland, football field, forest, industrial area, meadow, mountain, park, parking, pond, port, railway station, residential area, river, and viaduct. The dataset consists of a total of 1,005 images and each image has the size of 600 × 600 pixels