Technische Universität Braunschweig

UC Merced land use data

Arjun Arora, Usul Hamdi

Dr.-Ing. Mehdi Maboudi, Dr. Diaz Pedro

Introduction

1. This is a 21 class land use image dataset used for research purposes.

2. It contains 100 images for each of the 21 classes.

3. Each image has a resolution of 256x256 pixels.

4. The pixel resolution of each image is 0.3048 m.

5. The images are manually extracted form large USGS National map urban area Imagery collection for various urban areas around the country.

6. Classification of remote sensing image dataset.

Source

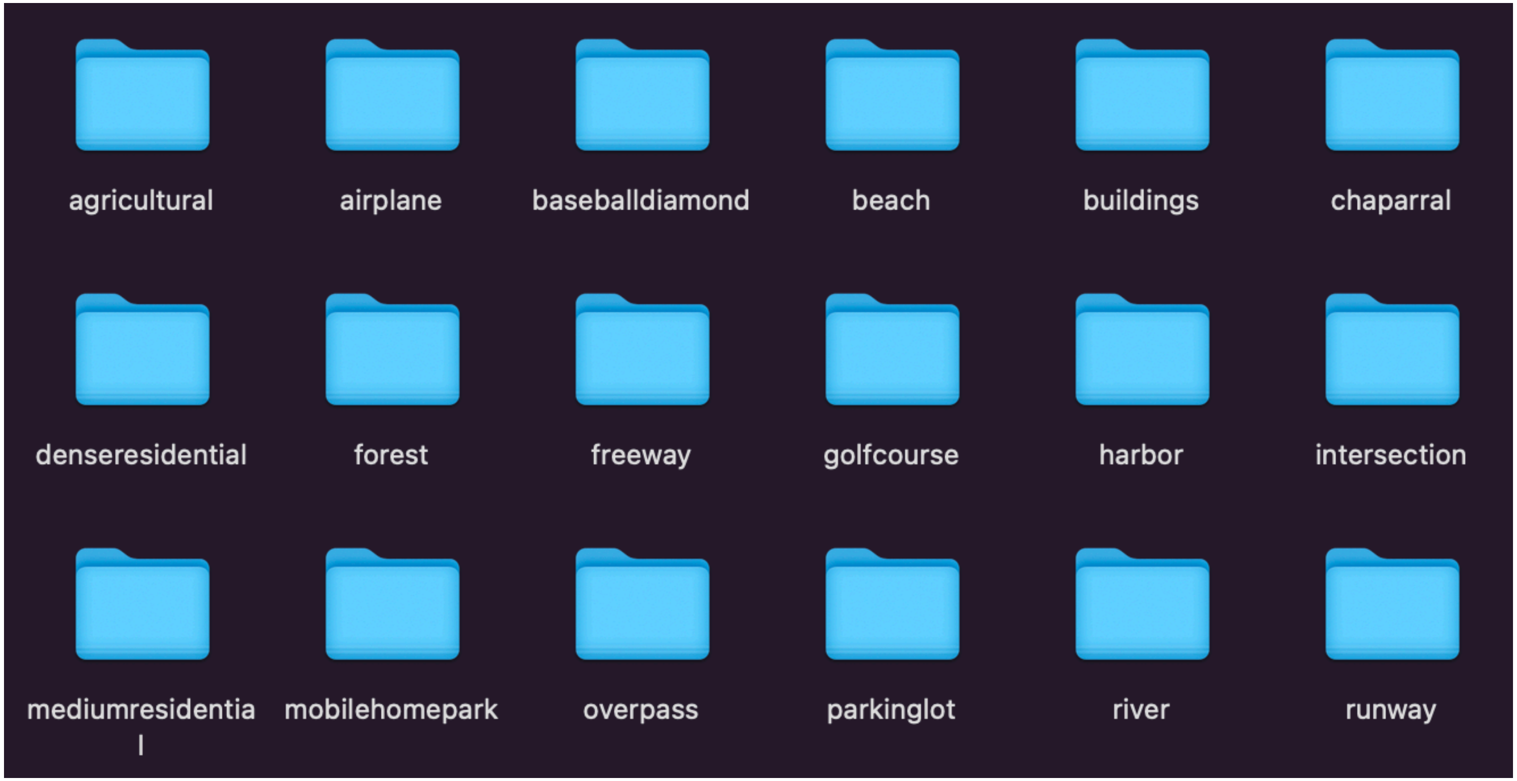
The data is from a 2010 paper by Yi Yang and Shawn Newsam called "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification," ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010.

Classes in dataset

There are 100 images for each of the following classes:

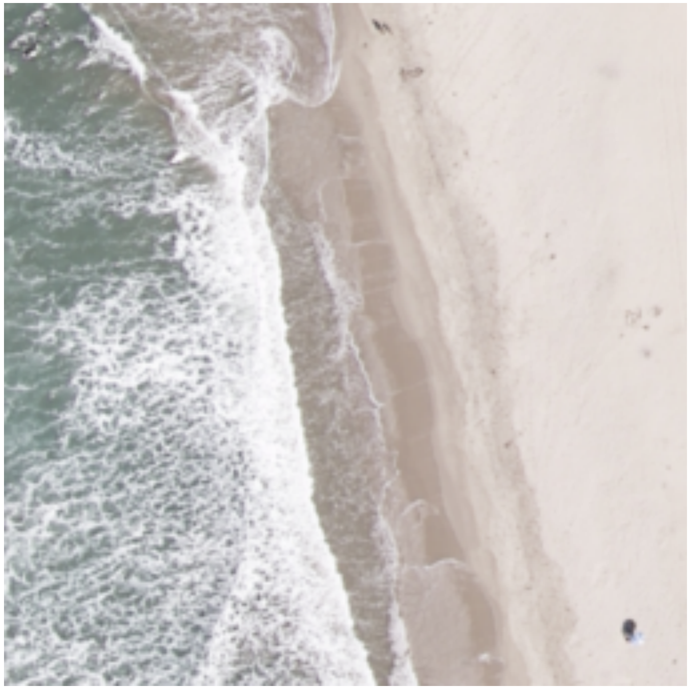
- agricultural
- airplane
- baseballdiamond
- beach
- buildings
- chaparral
- denseresidential
- forest
- freeway
- golfcourse
- harbor
- intersection
- mediumresidential
- mobilehomepark
- overpass
- parkinglot
- river
- runway
- sparseresidential
- storagetanks
- tenniscourt

Dataset raw



21 subfolders with class names

Data sample



Land use

The aim of this dataset is masking of regions where land is being used in satellite images obtained through Landsat.

Land use is an important problem of modern world which involves planning, Management and modifications of natural environment and available land.

It gives us a glimpse of our land use coverage, which type of establishments Cover the most space etc.

This is a basic dataset through which we can look at the approach to work on Land use data in remote sensing.

Dataset	UCM
Resolution	0.3048m
Classes	21
Total samples	2100
Obtained	Aerial imagery
Geolocations	USA

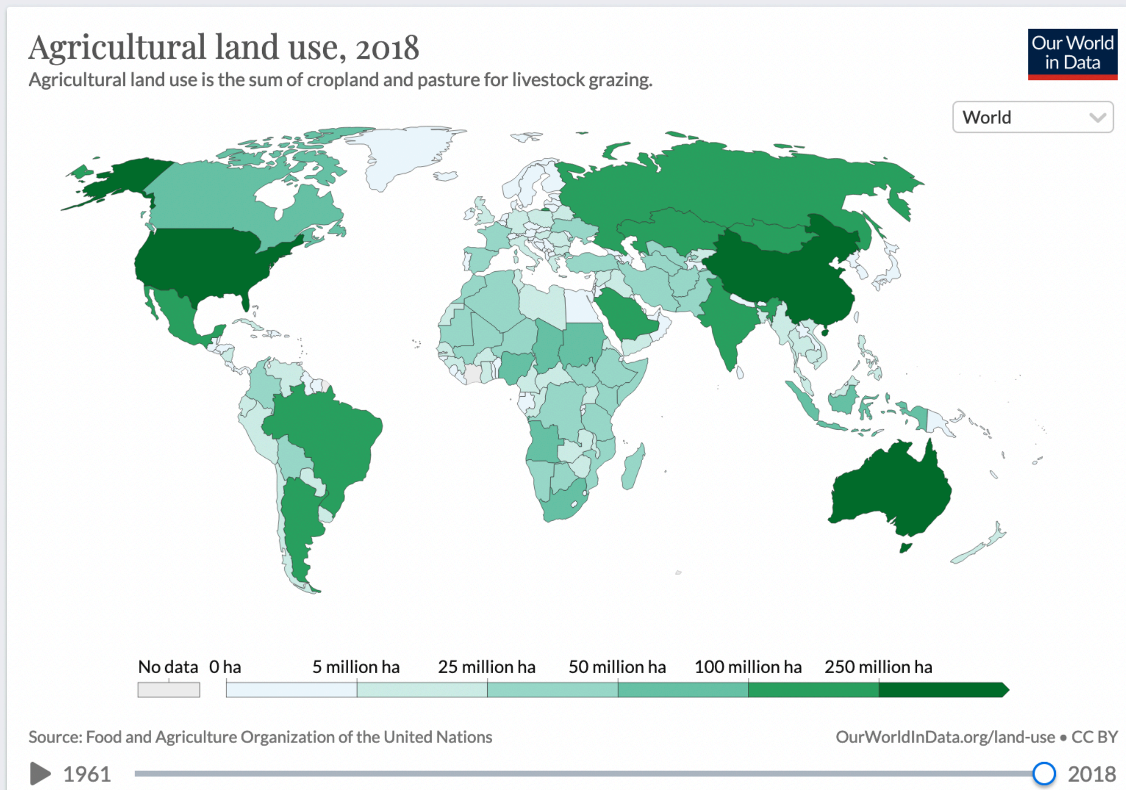
Data applications

A land use map of Europe—major non-natural land uses include arable farmland (yellow) and pasture (light green).

Data applications

US land use (2017)[11]

Use	acreage (M)	km^2 (M)	% of total
Pasture/range	654	2.647	35
Forest	538.6	2.18	28
Cropland	391.5	1.584	21
Special use*	168.8	0.683	9
Miscellaneous*	68.9	0.279	4
Urban	69.4	0.281	4
Total**	1,891	7.653	100



Data loading

1. The data is provided as a zip-file of 21 folders containing 100 images for each class.

2. We split the data to training, test and validation using a custom function.

3. We use keras.utils.image_dataset_from_directory function to then load the training test and validation data for each class at once.

4. This function returns a tf.data.Dataset type of tensorflow object containing the images.

5. As this function does not support tif image format("Tagged Image File Format", a high quality graphics image), we need to convert the images to jpg first.

If your directory structure is:

```
main_directory/
...class_a/
......a_image_1.jpg
......a_image_2.jpg
...class_b/
......b_image_1.jpg
......b_image_2.jpg
```

Then calling `image_dataset_from_directory(main_directory, labels='inferred')` will return a `tf.data.Dataset` that yields batches of images from the subdirectories `class_a` and `class_b`, together with labels 0 and 1 (0 corresponding to `class_a` and 1 corresponding to `class_b`).

Supported image formats: jpeg, png, bmp, gif. Animated gifs are truncated to the first frame.

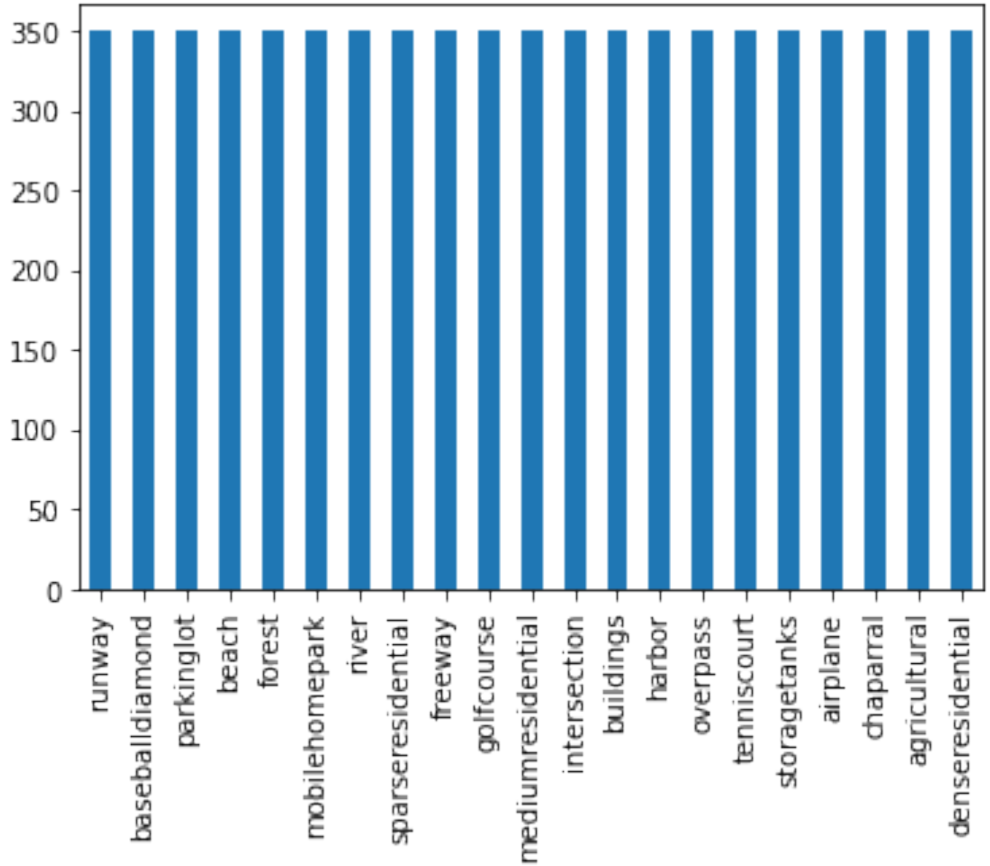
EDA - Classes

1. As each class has 100 samples, the classes are perfectly balanced.

2. We can use data augmentation to increase the image samples by using rotation, flipping, cropping etc.

3. In case of training with data augmentation, using a CNN model would be beneficial as its performance does not vary due to augmented images.

```
# Read train and test data from image directories
train = keras.utils.image_dataset_from_directory("archive/images_train_test_val/train")
test = keras.utils.image_dataset_from_directory("archive/images_train_test_val/test")
val = keras.utils.image_dataset_from_directory("archive/images_train_test_val/validation")
```



Samples per class

Assigning class numbers

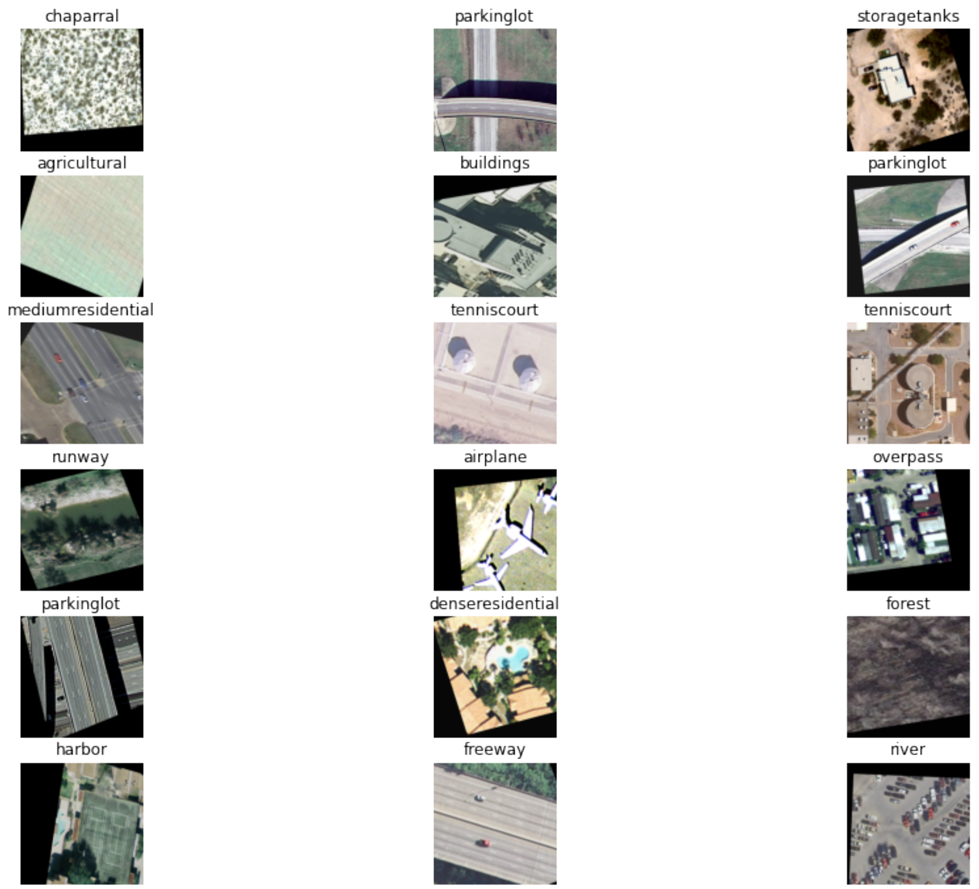
```
class_names = train.class_names
import json
with open("archive/label_map.json","r") as file:
    class_name_binarized = json.load(file)

num_classes = len(class_name_binarized)
class_names = list(class_name_binarized.keys())
class_name_binarized
```

[13]

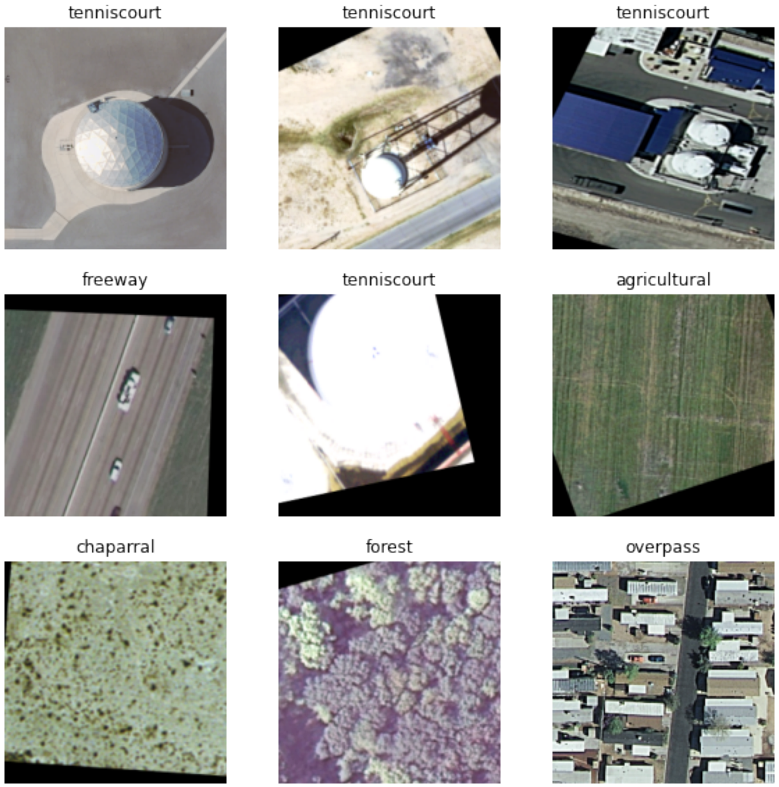
```
{'agricultural': 0,
 'airplane': 1,
 'baseballdiamond': 2,
 'beach': 3,
 'buildings': 4,
 'chaparral': 5,
 'denseresidential': 6,
 'forest': 7,
 'freeway': 8,
 'golfcourse': 9,
 'intersection': 10,
```

Train data



Training samples

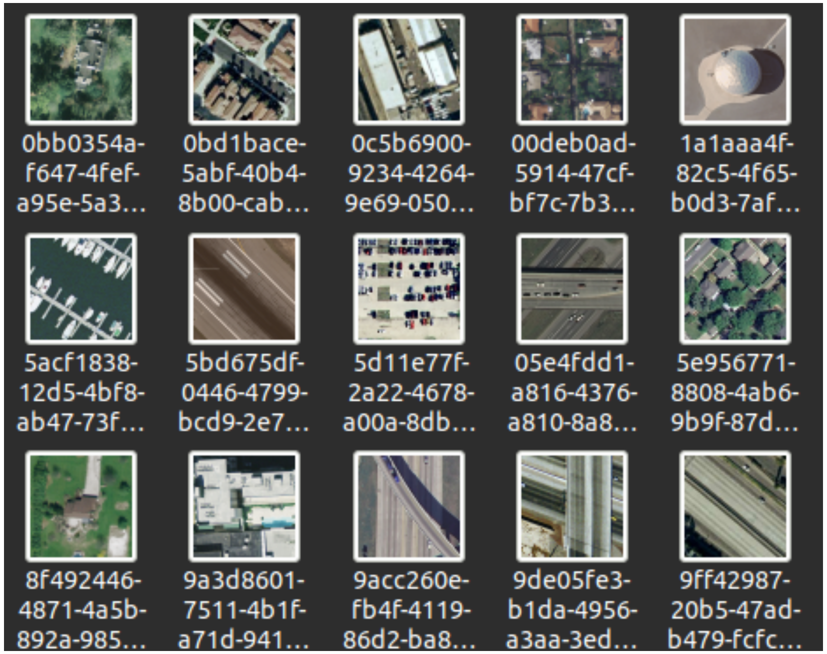
Test



Test samples

Human level performance

- After analyzing the dataset, it was clear that some classes (buildings, denseresidential, mediumresidential, mobilehomepark) were too close to each other.
- We wanted to analyze human level performance, we randomly sampled 10 image from the each class and put it into a folder with no mention of the label



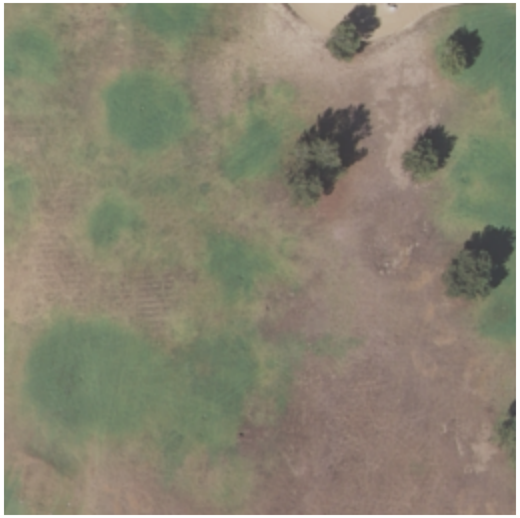
Human level performance

- And we tried to guess those labels ourselves.
- Out of 199 labels, we guessed 181 correctly.
- The ones that we classified incorrectly:

Label	Guess	Label	Guess
denseresidential	mobilehomepark	mobilehomepark	buildings
buildings	denseresidentia	golfcourse	agricultural
golfcourse	agricultural	mobilehomepark	denseresidential
intersection	overpass	mediumresidential	denseresidential
mediumresidential	denseresidential	sparseresidential	buildings
denseresidential	mobilehomepark	mediumresidential	denseresidential
mobilehomepark	denseresidential	denseresidential	mediumresidential
golfcourse	agricultural	forest	chaparral
golfcourse	agricultural	denseresidential	buildings

Human level performance

- As we suspected, medium, dense, spare residential areas, buildings and mobile park classes are easy to confuse.
- But also we also confused golf course with agricultural areas and vice versa.



Golf course or agricultural?

Dense or medium?

References

Yi Yang and Shawn Newsam, "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification," ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010.

- https://www.kaggle.com/datasets/apollo2506/landuse-scene-classification
- https://keras.io/api/data_loading/image/
- https://www.paintshoppro.com/en/pages/tif-file/
- https://elib.dlr.de/130278/1/08899080.pdf
- https://en.wikipedia.org/wiki/File:Europe_land_use_map.png

THANK YOU