Assignment 2
Report

Name: Jonas Schrade
Student Number: 01/1080887
Course (Instructor): Deep Learning for Social Science (Giordano Di Marzo)

1 Introduction

The goal of assignment 2 is to build, train, and evaluate a Convolutional Neural Network (CNN) to classify land use patterns from satellite imagery across Europe, specifically the EuroSAT dataset.

2 Results

2.1 Data

The data used to train, validate, and test the CNN model are Sentinel-2 satellite images sourced from the EuroSAT dataset, as published by Helber et al. [2018]. This dataset comprises 27,000 labeled and geo-referenced images spanning 10 Land Use and Land Cover (LULC) classes. Each image is a 64×64 pixel, 3-channel RGB image. The dataset is randomly split into training and testing subsets using an 85/15 ratio. An overview of the class distribution within the training set and a sample of representative images are provided in Figure 1 and Figure 2, respectively.

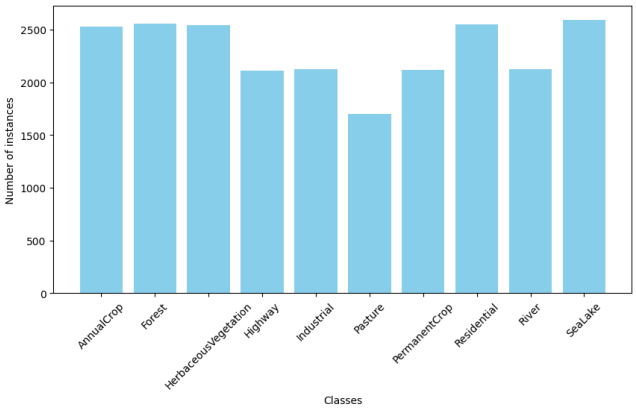

Figure 1: Target classes in training set.

Class frequencies range from approximately 1,500 to 2,500 images per class. Notably, the 'Pasture' class is somewhat underrepresented, followed by 'Highway', 'Industrial', 'Permanent Crop', and 'River'. Nevertheless, since all classes contain at least 1,500 instances, the risk of introducing significant class imbalance or bias during training is expected to be small.

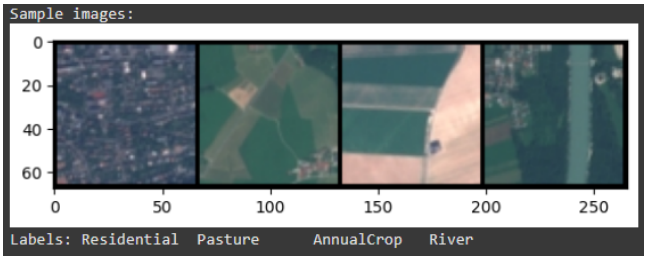

Figure 2: Sample images with labels from training set.

The training set is then split, so that training and validation represent 70% and 15% of the original data. To enhance generalization and reduce overfitting, augmentation is conducted only on the training set. This includes random horizontal flipping, random rotation, random affine, and color jitter (see notebook for visualization). Furthermore, image tensors in all three sets are normalized with mean and standard deviation of 0.5 across the three channels.

2.2 Model Architecture (see Figure 3)

The CNN architecture consists of six convolutional layers, each employing a 3×3 kernel with a padding of one pixel, followed by Rectified Linear Unit (ReLU) activation functions.

The first two convolutional layers transform the input from three channels to 32 feature maps. Each subsequent pair of convolutional layers doubles the number of feature maps, thereby increasing the representational capacity of the network. Max pooling operations with a 2×2 window are interleaved between each pair of convolutional layers, reducing the spatial dimensions by a factor of two. This downsampling decreases computational complexity and facilitates hierarchical feature extraction. To mitigate overfitting, dropout regularization is applied after each max pooling layer, randomly zeroing out 25% of the activations during training. The resulting feature maps, with 128 channels and spatial dimensions of 8×8, are flattened and passed through a fully connected layer with 64 neurons and ReLU activation. An additional dropout layer with a dropout rate of 50% is applied at this stage to further promote generalization. The final output layer is a linear classifier that generates logits for the 10 target classes. In total, the network contains 812,010 trainable parameters.

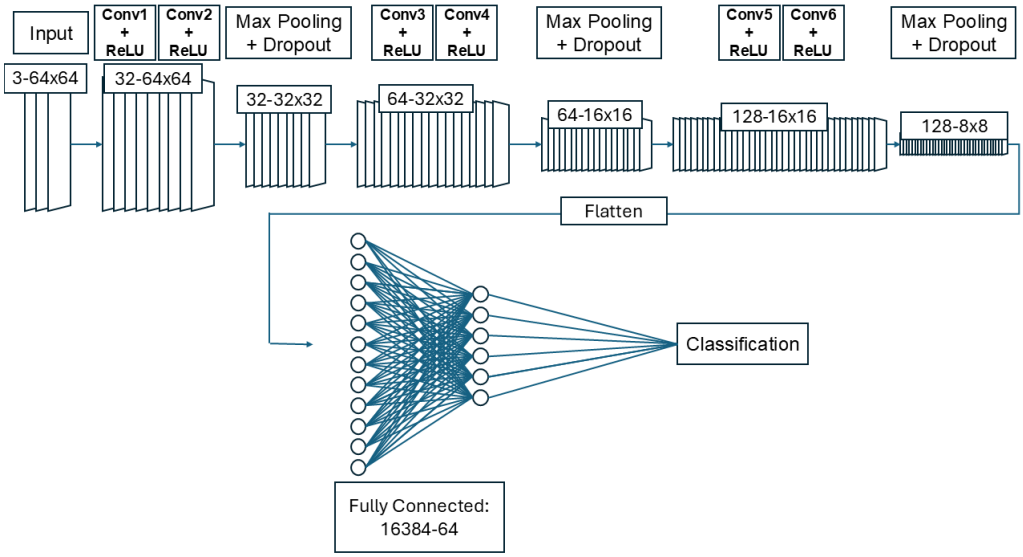

Figure 3: CNN Architecture.

2.3 Training

The CNN is trained using a standard gradient descent approach to minimize the loss function, specifically cross-entropy loss in this case. For model optimization, the AdamW algorithm is employed. Unlike the standard Adam optimizer, AdamW decouples weight decay from the gradient-based parameter updates, resulting in more consistent regularization and improved generalization performance. A batch size of 512 is chosen to accelerate training while maintaining stable gradient estimates. Both the learning rate and weight decay are set to 0.001 to ensure stable convergence and balanced regularization, thereby reducing the risk of overfitting. Training is configured to run for a maximum of 50 epochs; however, an early stopping criterion with a patience of 5 epochs is applied to halt training if the validation loss does not improve. The descent of training and validation loss is visualized in Figure 4. The training process exhibits a generally stable decline in both training and validation loss curves, interrupted only by minor fluctuations. Early stopping is triggered at epoch 47. Notably, from epoch 36 onward, the validation loss begins to exceed the training loss and subsequently plateaus, with only marginal improvements observed there-

after. The training loss ultimately reaches a minimum of 0.217, while the validation loss achieves its lowest value of 0.272 at epoch 42, before increasing slightly to 0.283 by the end of training.

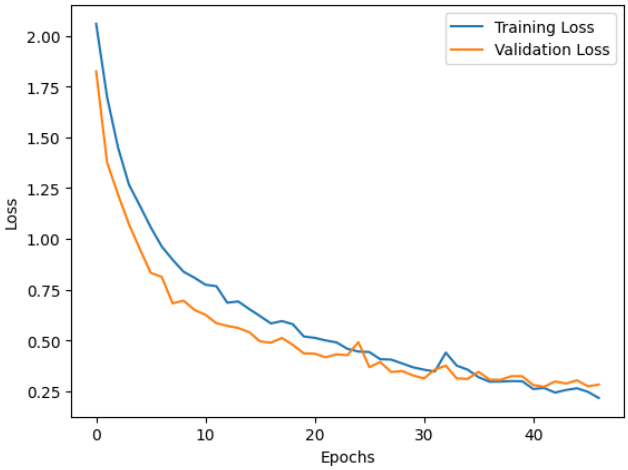

Figure 4: Learning curves of CNN training.

Table 1: Classification Report.

Class	Precision	Recall	F1-score
AnnualCrop	0.906	0.960	0.932
Forest	0.971	0.975	0.973
HerbaceousVegetation	0.925	0.882	0.903
Highway	0.930	0.913	0.921
Industrial	0.989	0.960	0.975
Pasture	0.933	0.933	0.933
PermanentCrop	0.892	0.934	0.912
Residential	0.967	0.982	0.975
River	0.963	0.912	0.937
SeaLake	0.981	0.990	0.985
Accuracy			0.945
Macro Avg	0.946	0.944	0.945
Weighted Avg	0.945	0.945	0.945

2.4 Evaluation

The CNN model achieves an image classification accuracy of 94.5% on the test set. Precision and recall show similarly high performance, with macro-averaged values indicating consistent results across classes (see Table 1). The LULC categories 'SeaLake', 'Industrial', 'Residential', and 'Forest' achieve the highest F1-scores, especially driven by their high recall values. This performance likely stems from the distinct and easily recognizable visual characteristics these classes exhibit in satellite imagery. In contrast, classes such as 'HerbaceousVegetation', 'PermanentCrop', 'AnnualCrop', and 'Pasture', along with 'Highway' and 'River', show comparatively lower F1-scores, reflecting somewhat reduced precision and recall. The lower performance among agricultural categories may be attributed to their visual similarity and frequent misclassification among such land uses. Additionally, the linear and spatially analogous features of highways and rivers may contribute to their confusion with each other and their surrounding environment. The confusion matrix offers further insight into the model's classification behavior and common misclassifications.

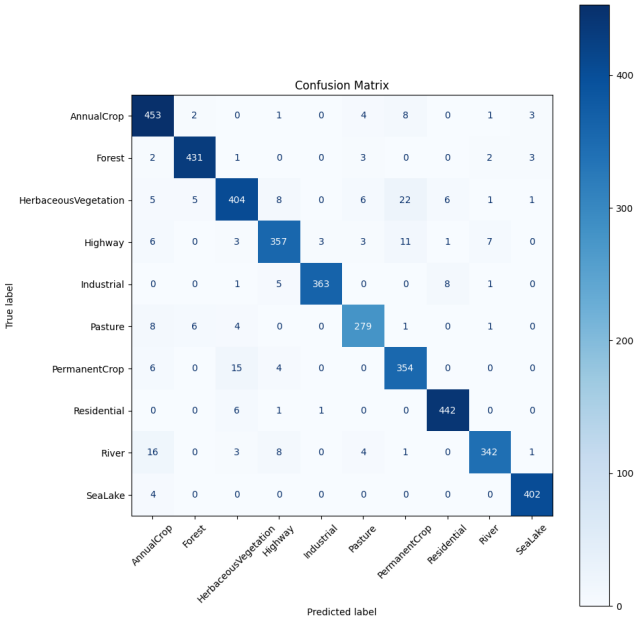

Figure 5: Confusion Matrix CNN.

3 Conclusion

This report outlines the development and training of a CNN for land use classification using the EuroSAT satellite image dataset. The methodology employed standard procedures for data preprocessing, augmentation, model configuration, training, and evaluation. Overall, the results indicate that the model effectively distinguishes between various land cover types based on the satellite imagery. Nonetheless, further performance enhancements may be warranted for the classification of visually similar land cover classes—such as highways and rivers or different types of agricultural land—which the model sporadically struggles to differentiate.

4 Bonus

4.1 MLP vs. CNN

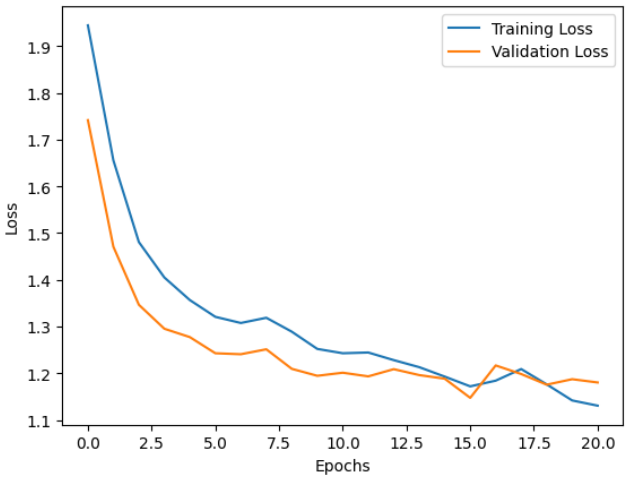

Figure 6: Learning curves of MLP training.

A standard multi-layer perceptron (MLP) was configured and trained for image classification. The model architecture begins with an input layer that accepts flattened 3-channel images of size 64×64, which are fed into a dense layer of 512 neurons, followed by a ReLU activation function and a 50% dropout rate. The subsequent layer reduces the dimensionality to 256 neurons, again followed by ReLU activation and a 30% dropout. The final hidden layer consists of 128 neurons with ReLU activation, but no dropout is applied. The output layer maps the representation to 10 classes for classification. In sum, the model architecture comprises 6,457,482 trainable parameters, significantly more than the CNN. The hyperparameters, loss criterion, and optimization algorithm for the MLP were configured identically to those used for the CNN.

As shown in Figure 6, training was halted after epoch 21 due to the early stopping criterion, with a training loss of 1.131 and a validation loss of 1.181. Notably, the minimum validation loss of 1.146 was achieved at epoch 16, suggesting that while the training loss continued to decline, the validation loss plateaued, signaling the onset of overfitting.

Table 2: Classification Report (MLP).

Class	Precision	Recall	F1-score
AnnualCrop	0.603	0.663	0.632
Forest	0.797	0.824	0.810
HerbaceousVegetation	0.616	0.476	0.537
Highway	0.546	0.200	0.292
Industrial	0.881	0.862	0.872
Pasture	0.573	0.813	0.672
PermanentCrop	0.641	0.509	0.568
Residential	0.514	0.796	0.625
River	0.549	0.653	0.597
SeaLake	0.847	0.709	0.772
Accuracy			0.648
Macro Avg	0.657	0.651	0.638
Weighted Avg	0.657	0.648	0.637

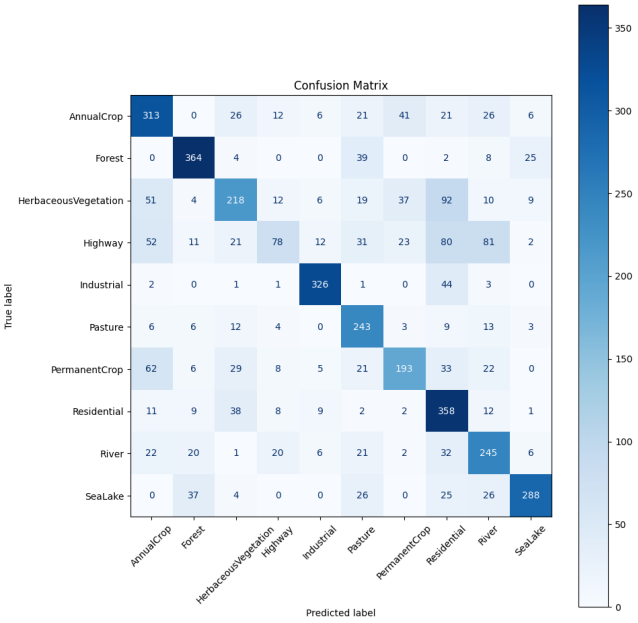

Figure 7: Confusion Matrix MLP.

While the CNN demonstrated robust classification performance, the MLP achieved only moderate results, with an overall accuracy of 65.8%. Table 2 summarizes class-specific precision, recall, and F1-scores. Notably, performance for classes such as 'Highway' and 'HerbaceousVegetation' remained low. Figure 7 illustrates common misclassifications, particularly the tendency to label diverse patterns as 'Residential'. Figure 8 shows how both models tend to struggle in distinguishing some agricultural land use and linear spatial features like highways and rivers.

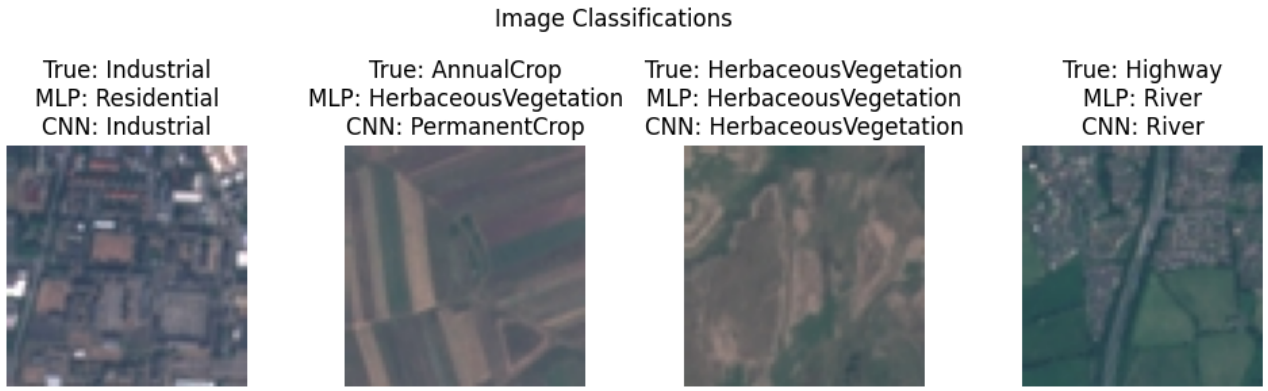

Figure 8: Classification Examples.

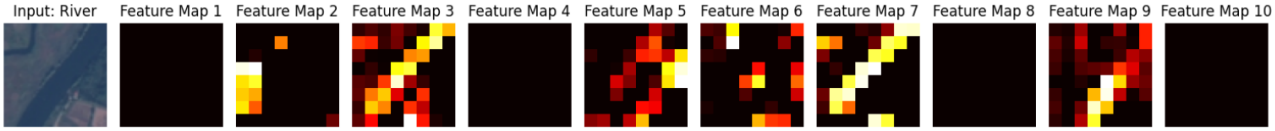

Figure 9: Extracted feature maps of last convolutional layer (example).

4.2 Features

An illustrative example of feature maps and corresponding activations from the final convolutional layer is presented in Figure 9. In this case, the model successfully captures the distinctive spatial characteristics of a river and its surrounding landscape, effectively delineating the dark, elongated structure from its environment. The observed activations demonstrate the model's capacity to detect salient visual cues such as edges, textures, and contrast transitions typically associated with (long) water bodies. Notably, while some filters exhibit pronounced activation, several other filters remain largely inactive, indicating that the most critical features are extracted by a limited subset of filters within this layer.

References

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. https://doi.org/10.5281/zenodo.7711810, 2018. Dataset published on Zenodo. Appears in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 12, No. 7, pp. 2217–2226.