Distinguishing Urban versus Non-Urban Land Masses to Predict Urban Expansion using EuroSAT Satellite Imagery

Samuel Nelson
University of Minnesota-Twin Cities

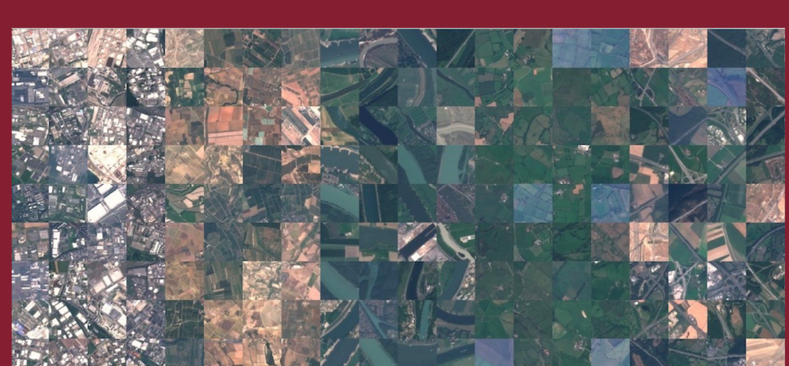
Introduction

Satellite Imagery

Satellite imagery provides spatial data to observe Earth's surface from above, and the collection and processing of this data contribute to the advancement of global Geographical Information Systems (GIS). The development of dynamic GIS is essential for innovating how we interact with our environment and surroundings.

The use of satellite imagery is especially prevalent within in the study of Urban Expansion (UE)- a major global phenomenon impacting infrastructure, the environment, and resource allocation. Urban planners rely on critical information including the identification and categorization of diverse land cover and land use to make informed decisions and guide infrastructure development to suit Earth's rapidly growing population.

The EuroSAT dataset, derived from the European Space Agency's Sentinel-2 mission, offers a plethora of labeled satellite images for 10 categories of land cover across regions of Europe. EuroSAT is widely used for land cover classification. A state-of-the-art convolutional neural network (CNN) benchmark on this dataset reached an overall accuracy of 98.57% for the original 10-class classification task.



Project Scope

This project reframes that task into a binary classification problem, predicting Urban versus Non-Urban land cover to evaluate whether machine learning models can effectively distinguish urban areas by using only RGB satellite images.

To support this analysis, I implemented a 3-layer CNN and compared its performance with a random forest classifier benchmark. Both models were trained with EuroSAT RGB satellite imagery to classify each image as Urban or Non-Urban. Urban categories include the Highway, Industrial, and Residential class labels. Non-Urban categories include the AnnualCrop, Forest, HerbaceousVegetation, Pasture, PermanentCrop, River, and SeaLake class labels. Performance was measured using classification accuracy, precision, recall, F1-score, and AUC-ROC on a held-out test set.

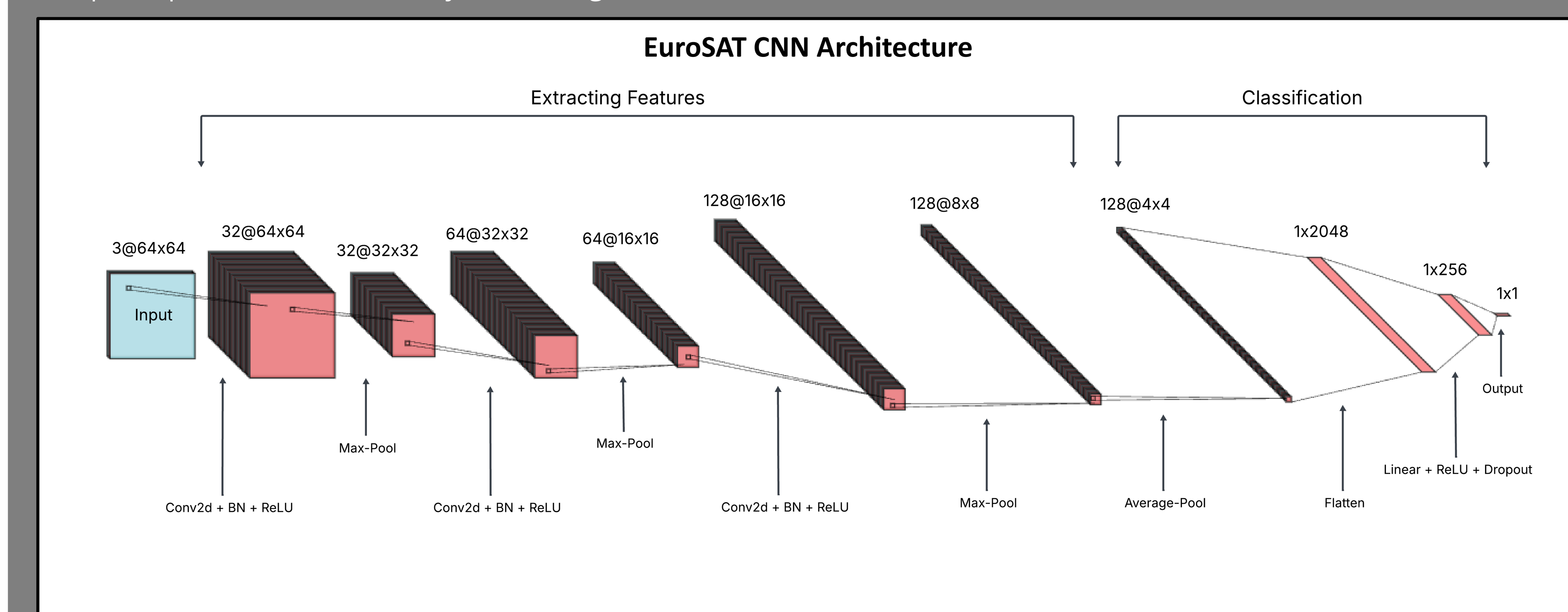
Methodology

Dataset

The EuroSAT RGB subset was used, containing 27,000 satellite images (64x64 pixels, 3 channels). The original 10 classes were re-labeled into binary classes: Urban (Highway, Industrial, Residential) versus Non-Urban (remaining classes). Urban images comprised ~22% of the dataset. To address class imbalance, I applied random down sampling of the Non-Urban class in the training set. Both models utilized a 70% training, 15% validation, and 15% testing split to ensure equal class distribution and promote a consistent approach to comparing the effectiveness of each.

3-Layer Convolutional Neural Network (CNN)

- **Architecture**: Comprised of 3 convolutional blocks, each consisting of a convolution layer (Conv2d), batch normalization (BN), rectified linear unit (ReLU) activation, and max pooling. This structure extracts hierarchical spatial features (distance, direction, shape) from the original input images.
- Followed by adaptive average pooling which standardizes spatial dimensions to output a 4x4 feature map, regardless of input image size.
- The final classification is done through two fully-connected network layers, with regularization enabled via dropout (p=0.5), outputting a single logit for binary classification (Urban versus Non-Urban).
- **Training Details**:
  - **Loss Function**: Binary Cross Entropy with Logits (BCELossWithLogits)
  - **Optimizer**: Adam
  - **Batch Size**: 64
  - **Learning Rate**: 0.0001
  - **Epochs**: 20
- **Rationale**: CNNs are ideal for spatially structured data like satellite imagery. The architecture is both simple and performant. It balances model capacity and efficient training time, allowing the model to learn low to high-level spatial patterns with relatively small image resolutions.



Random Forest (RF) Classifier

- **Input Representation**: Features were extracted by flattening each 3x64x64 input image into a 12,288-dimensional vector, removing spatial structure but retaining color and texture information.
- **Training Details**:
  - **Number of Trees**: 100
  - **Criterion**: Gini impurity
- **Rationale**: Serves as an interpretable, robust non-deep learning baseline to evaluate the difference in performance compared with CNNs that utilize spatial feature extraction.

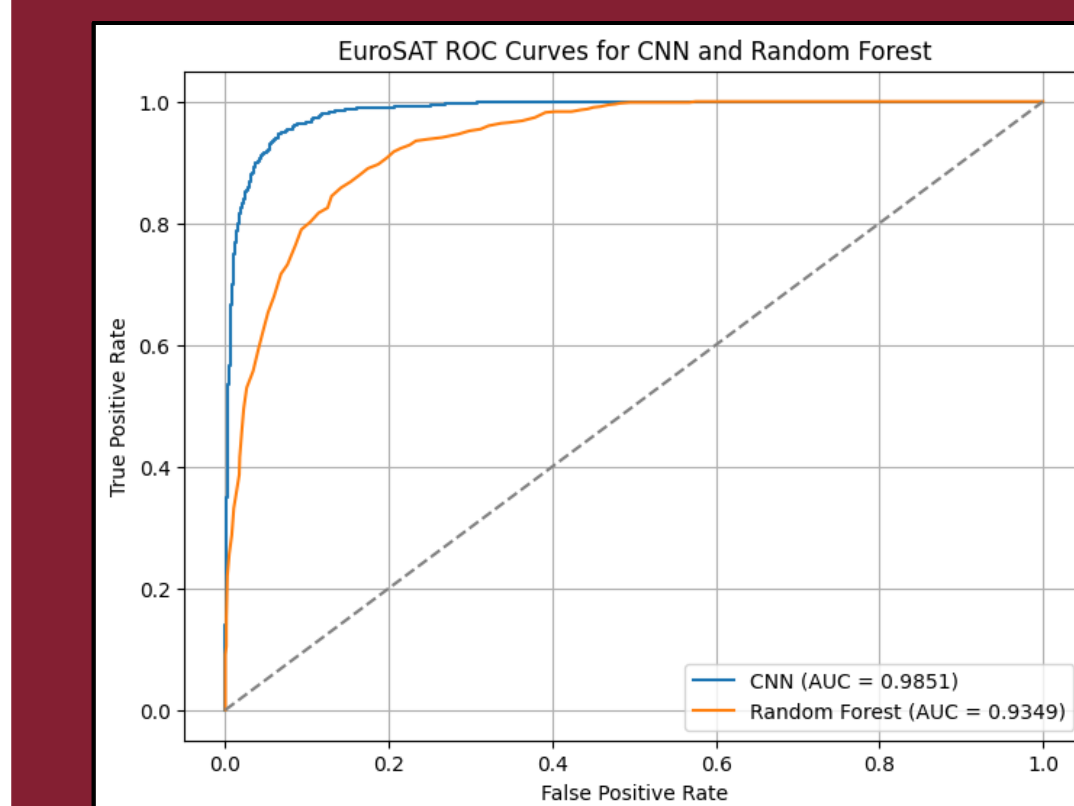
Results & Discussion

AUC-ROC Curves

- CNN outperformed random forest on the binary classification task.
  - **CNN AUC**: 0.9851
  - **RF AUC**: 0.9349
- The CNN's ROC curve illustrates a steeper climb toward the top-left corner, indicating better separation between Urban and Non-Urban classes.

Performance Metrics

- **Urban Class**: CNN had higher precision (0.87) and F1-score (0.90), suggesting better handling of false positive and overall balance. RF had higher recall (0.91), suggesting it captured more actual urban examples but at the cost of more false positives.
- **Non-Urban Class**: CNN led in all metrics (particularly with an F1-score of 0.95), suggesting strong performance across the larger class.
- **Overall Accuracy**: CNN had an accuracy of 0.94 while RF had an accuracy of 0.83, indicating the CNN's strong advantage in spatial feature extraction and model capacity.



Urban		
	CNN	RF
Precision	0.87	0.67
Recall	0.94	0.91
F1-Score	0.90	0.77

Non-Urban		
	CNN	RF
Precision	0.97	0.95
Recall	0.94	0.80
F1-Score	0.95	0.87

Overall		
	CNN	RF
Accuracy	0.94	0.83

Discussion

The CNN had an edge over the RF classifier because it can handle features more effectively due to convolutional layers, and the shallow architecture was still expressive enough due to the balanced dataset and relatively small image size. The RF classifier was limited by its flattened input (removing spatial context), likely affecting its separation of Urban and Non-Urban classes. Misclassifications could be a result of ambiguous regions, such as bodies of water near urban areas. Future rework could include refining class labeling criteria using geospatial context to ensure class boundaries and reduce overlap, while incorporating multispectral data to provide richer context.

References

[1] Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. Patrick Helber, Benjamin Bischke, Andreas Dengel, Damian Borth. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019.

[2] Introducing EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. Patrick Helber, Benjamin Bischke, Andreas Dengel. 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018.