

Project Proposal Template

EE 541: A Computational Introduction to Deep Learning

Project Title: Satellite Land Use Classification

Group Members: [Student Name 1, Student Name 2]

Selected Topic: [Which of the four project topics you selected]

Problem Description

We will build a deep learning system to classify satellite imagery into land use categories. The task involves taking overhead satellite images and predicting which type of land use they represent—residential areas, forests, agricultural fields, industrial zones, or bodies of water. This is a multi-class image classification problem where we must distinguish between visually similar categories using only RGB or multispectral satellite data.

Satellite images differ substantially from typical photographs used in image classification benchmarks. The overhead perspective removes familiar visual cues like object orientation and relative size. Resolution is lower than natural photographs, making fine details difficult to distinguish. Different land use types can appear visually similar—residential and commercial areas both show dense building patterns, while different crop types may look nearly identical. Seasonal variation affects vegetation appearance, atmospheric conditions introduce noise and color shifts, and the spectral characteristics of materials become important discriminators that aren't visible in standard RGB imagery.

The dataset we will use is EuroSAT, containing 27,000 labeled images from Sentinel-2 satellites covering 10 land use classes. Each image is 64×64 pixels captured at 10-meter spatial resolution across Europe. The dataset provides both RGB and multispectral bands (13 total), allowing us to explore whether near-infrared and shortwave infrared bands improve classification over visible light alone. Classes are balanced with 2,700 examples each, simplifying initial experiments but not reflecting real-world [...]



Figure 1: Example satellite images from different land use classes. Top row shows residential (dense housing patterns), forest (vegetation texture), and annual crop (regular field geometry). Bottom row shows highway (linear features), industrial (large buildings and pavement), and sea/lake (water signature). Note that at 64×64 resolution, distinguishing features often rely on texture and patterns rather than individual objects.

Your problem description should explain the task in your own words, what makes it challenging from a deep learning perspective, and key characteristics of the dataset. Demonstrate understanding of the specific topic you selected.

Dataset Analysis

The EuroSAT dataset contains 27,000 labeled satellite images across 10 land use categories: annual crop, forest, herbaceous vegetation, highway, industrial, pasture, permanent crop, residential, river, and sea/lake. Images are 64×64 pixel RGB crops extracted from Sentinel-2 satellite scenes captured between 2015-2017 over various European locations. The dataset provides 13 spectral bands total—visible RGB plus near-infrared and shortwave infrared bands that capture information invisible to human eyes.

Class distribution is balanced at 2,700 images per class. While this simplifies initial modeling by avoiding class imbalance issues, it doesn't reflect real-world land use distributions where some categories like industrial zones are much rarer than forests or agricultural areas. The balanced distribution means accuracy is a reasonable primary metric, though we'll also examine per-class performance to understand which categories are harder to distinguish.

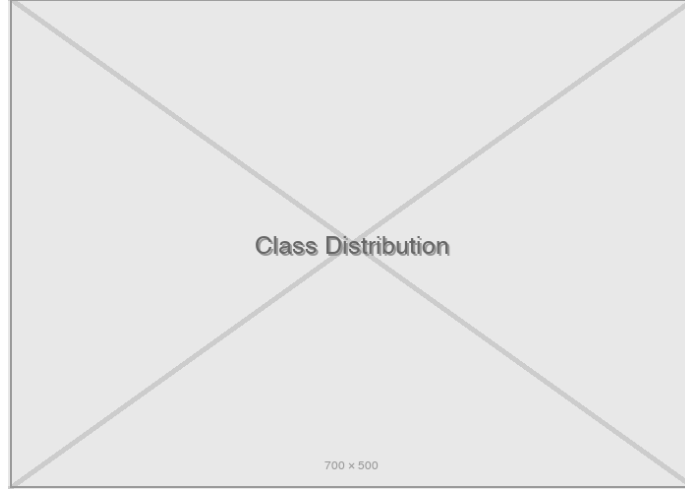


Figure 2: Distribution of training samples across 10 land use categories after 80/20 train/test split. Each class contains 2,160 training images and 540 test images. The balanced distribution eliminates class imbalance as a variable in our experiments.

Spatial resolution is 10 meters per pixel, meaning each 64×64 image covers a 640×640 meter area on the ground. This resolution is sufficient to see large structures and land patterns but too coarse for identifying individual cars or small buildings. Residential areas appear as textured regions of buildings rather than distinct houses. Highways show as linear features but individual lanes aren't visible. Forest and agricultural areas show characteristic textures from vegetation patterns.

The multispectral bands beyond RGB provide information about vegetation health, water content, and material composition. Near-infrared bands (Band 8 at 842nm) are strongly reflected by healthy vegetation but absorbed by water, making them valuable for distinguishing vegetated areas from water bodies. Short-wave infrared bands (Bands 11-12 at 1610nm and 2190nm) help differentiate soil types and measure moisture content. Previous work on this dataset achieved 98.6% accuracy using all 13 bands versus 98.2% with RGB only, suggesting that most discriminative information is visible in RGB but multispectral bands [...]

Your dataset analysis should characterize the data (size, classes, dimensions, splits), include exploratory visualizations, discuss what makes classification challenging, and analyze dataset properties that will affect your modeling choices.

Literature Survey

Convolutional neural networks have become the standard approach for image classification since Krizhevsky et al. (2012) demonstrated that deep CNNs with multiple convolutional layers substantially outperform traditional computer vision methods. AlexNet introduced key architectural ideas still used today—stacking convolutional layers to learn hierarchical features, using ReLU activations to enable training of deep networks, and applying dropout to prevent overfitting. The core insight that convolutional filters can automatically learn relevant features from data rather than requiring hand-designed feature extractors like SIFT or HOG applies directly to satellite imagery classification.

Architectural depth matters for learning complex visual patterns. Simonyan & Zisserman (2015) showed with VGGNet that using many layers with small (3×3) convolutional filters outperforms

fewer layers with larger filters. For satellite imagery where discriminative features include both fine textures and large-scale spatial patterns, deep networks that can capture multiple scales of information should perform better than shallow networks. He et al. (2016) introduced residual connections in ResNet, enabling training of very deep networks (50+ layers) by allowing gradients to flow through skip connections. Whether such depth is necessary for our 64×64 images at 10-meter resolution—where there’s limited fine detail to capture—is a question we’ll [...]

Your literature survey should include 5-8 relevant sources that inform your approach. For each source, explain what it contributes to your understanding and how it guides your planned experiments. This is focused research demonstrating you understand the context, not an exhaustive review.

Proposed Approach

We plan to investigate several architectures of increasing complexity to understand what level of model capacity is necessary for this task. Our baseline model will use a straightforward architecture with three convolutional blocks. The first block applies 32 filters of size 3×3 , reducing spatial dimensions from 64×64 to 32×32 through 2×2 max pooling. The second block doubles the filters to 64, reducing to 16×16 . The third block increases to 128 filters, reducing to 8×8 . After flattening, two fully connected layers with 256 and 10 units produce class logits.

We will then implement a VGG-style architecture using 3×3 convolutions exclusively, stacked in groups with increasing filter counts (64, 128, 256, 512). For transfer learning experiments, we’ll use pretrained ResNet-18 initialized with ImageNet weights, comparing frozen features versus fine-tuning to understand how much adaptation to satellite imagery is [...]

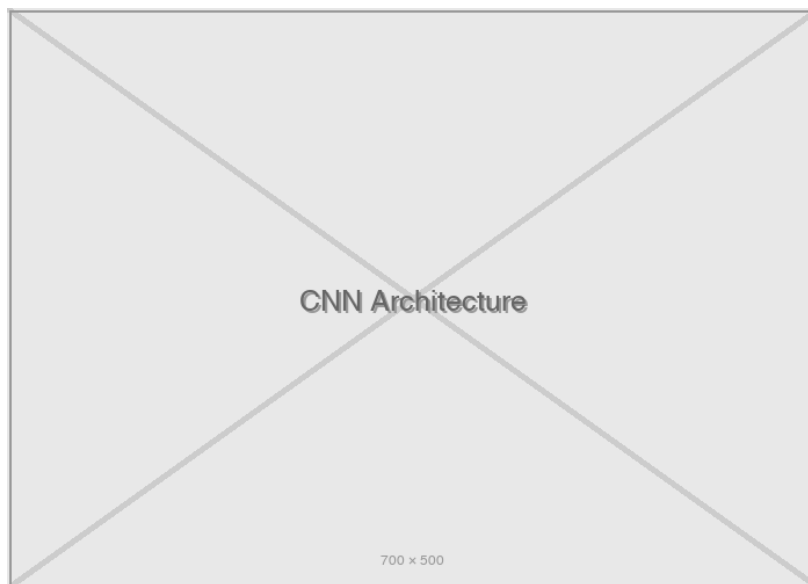


Figure 3: Architecture diagram showing the progression from simple baseline CNN through VGG-style deep network to ResNet with skip connections. Each architecture shows the sequence of convolutional layers, feature map dimensions, and fully connected classification head.

Data preprocessing will normalize images to $[0,1]$ range by dividing pixel values by 255. For pretrained models expecting ImageNet normalization, we’ll apply those transformations. We will

apply data augmentation during training including random horizontal and vertical flips (probability 0.5 each), random 90-degree rotations, and small random crops with padding.



Figure 4: Preprocessing pipeline showing raw satellite image converted to tensor, normalized to $[0,1]$ range, optionally standardized with ImageNet statistics for pretrained models, and augmented during training with flips, rotations, and color jittering.

We'll split the data 80/15/5 into training, validation, and test sets (21,600 train, 4,050 validation, 1,350 test).

Loss Function: We will minimize cross-entropy loss for this multi-class classification problem:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c})$$

where N is batch size, $C = 10$ classes, $y_{i,c}$ is the one-hot encoded true label, and $\hat{y}_{i,c}$ is the predicted probability from softmax. This loss penalizes confident wrong predictions more than uncertain ones.

Evaluation Metrics: Primary metric is classification accuracy:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

We'll also compute per-class precision and recall. For class c :

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}, \quad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c}$$

where TP_c is true positives for class c . These metrics reveal which classes are [...]

Your proposed approach should describe the architectures you'll explore, data preprocessing and augmentation strategies, training procedures, and evaluation methodology including specific loss functions and metrics with equations. Show systematic experimental planning based on your literature survey. This is your initial plan—it will evolve during experimentation.

Timeline and Division of Work

- **15-21 Jan:** Download dataset, implement data loading and preprocessing (Member 1), implement baseline CNN and training loop (Member 2), verify pipeline on small subset
- **22-28 Jan:** Train baseline models, implement VGG-style architecture (Member 1), implement data augmentation (Member 1), begin transfer learning with pretrained ResNet (Member 2)
- **29 Jan - 04 Feb:** Complete architecture comparisons, run ablation studies on augmentation and depth (Member 1), experiment with fine-tuning strategies (Member 2)
- **05-11 Feb:** Finalize best model, analyze results and failure cases (both), generate visualizations (Member 1), draft report (Member 2)
- **12-18 Feb:** Complete report with analysis (both), finalize code documentation (Member 1), prepare model card [...]

Your timeline should show realistic weekly milestones and division of responsibilities. Demonstrate that you've planned the work systematically and both members will contribute substantially.

References

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *NeurIPS 2012*.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *ICLR 2015*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *CVPR 2016*.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *NeurIPS 2014*.

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint*.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*.

Your references should include papers that informed your approach. Use standard citation format with authors, title, venue, and year.