TerraVision: Benchmarking Land Use and Land Cover with Deep Learning

1st Samiha Muntaha Mahin
International Islamic University
Chittagong, Bangladesh
samihamahin36@gmail.com

2nd Tahmina Hasan
International Islamic University
Chittagong, Bangladesh
tahminah587@gmail.com

3rd Rezowan Sifat
International Islamic University
Chittagong, Bangladesh
navidrezowansifat@gmail.com

4th Salman Shahriar
International Islamic University
Chittagong, Bangladesh
salmanshahriar16@gmail.com

5th Mahamudul Hasan Sunny
Mirpur College
Dhaka, Bangladesh
Mahamudul.editor@gmail.com

6th Sara Karim
International Islamic University
Chittagong, Bangladesh
sarabintekarim@gmail.com

***Abstract*—This paper addresses the problem of land use and land cover (LULC) classification using Sentinel-2 satellite imagery. We utilize the RGB subset of the EuroSAT dataset, encompassing 10 categories with a total of 5,000 labeled images, and perform data balancing and augmentation to improve model generalization. A convolutional neural network based on EfficientNetB0 is trained on this dataset, achieving a final test accuracy of 88.3%. The model demonstrates robust performance across diverse land cover classes, as validated by confusion matrices and classification reports. These results highlight the potential of deep learning models for practical Earth observation tasks, including monitoring land use changes and supporting geographic mapping. The processed dataset and code provide a reproducible framework for future remote sensing and machine learning research.**

***Index Terms*—Remote Sensing, Sentinel-2, Satellite Image Classification, Land Use and Land Cover, EfficientNet, Deep Learning, Dataset, Image Augmentation.**

I. Introduction

Satellite imagery has become an indispensable tool for monitoring Earth's surface. Programs like ESA's Copernicus and NASA's Landsat provide free and continuous access to high-quality imagery, enabling applications in agriculture, forestry, urban planning, environmental monitoring, and disaster assessment.

However, raw satellite images are not immediately usable—they require interpretation and labeling to transform pixel data into actionable information. A key step in this process is land use and land cover (LULC) classification, which categorizes regions as forest, water, urban, agricultural, or other land types.

Deep learning, especially convolutional neural networks (CNNs), has significantly advanced image classification tasks. Nevertheless, their performance depends heavily on well-prepared datasets. Many existing LULC datasets are either small, unbalanced, or focused on multi-spectral bands not suitable for standard RGB-based models. This limits the ability to train robust CNNs on realistic satellite imagery.

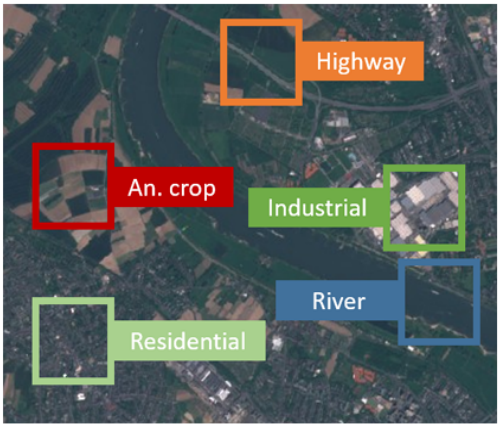

Fig. 1. illustrates sample RGB satellite images from the dataset, showing classes such as Industrial, Forest, River, Residential, Highway and Crops.

To address these challenges, we use the RGB subset of the EuroSAT dataset, which contains 10 LULC classes and 5,000 labeled images. We implement data balancing and augmentation strategies to improve class representation and generalization. Using EfficientNetB0, we achieve a final test accuracy of 88.3%, demonstrating strong classification performance across diverse land cover types.

Our contributions include:

- Preprocessing and balancing of the RGB EuroSAT dataset for robust CNN training.
- Implementation of an EfficientNetB0-based deep learning model for LULC classification.
- Evaluation of model performance using classification reports, confusion matrices, and accuracy metrics.
- A reproducible framework for RGB satellite image classification applicable to real-world monitoring tasks.

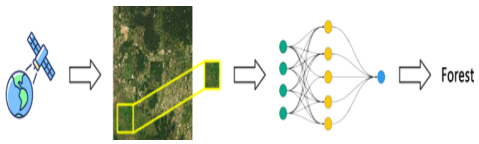

Fig. 2. This illustration depicts patch-based land use classification from satellite images. Extracted patches are fed into a neural network, which automatically labels the land type or usage.

II. Related Works

This section reviews prior research on land use and land cover (LULC) classification. We discuss both aerial and satellite image datasets used for this purpose, as well as current state-of-the-art classification techniques.

A. Classification Dataset

Classifying remotely sensed images remains a challenging task, largely due to the limited availability of reliably labeled ground truth datasets. One widely used and well-studied dataset is the UC Merced (UCM) land use dataset, introduced by Yang et al [1]. It includes 21 land use and land cover categories, with 100 images per category. Each image is 256×256 pixels with a spatial resolution of approximately 30 cm per pixel, in RGB format, and was captured from aircraft imagery in the USGS National Map Urban Area collection. However, the dataset's small size per class poses limitations for large-scale analysis.

To address this, several studies manually generated new datasets using high-resolution Google Earth images [2], [3], [4], [5]. Notable examples include PatternNet [6] and NWPU-RESISC45 [7]. These datasets feature very high spatial resolution (up to 30 cm per pixel), but the time-consuming labeling process means each class still contains only a few hundred images. The Aerial Image Dataset (AID) is one of the largest, containing 30 classes with 200–400 images per class, using 600×600 pixel images also from Google Earth.

Compared to the EuroSAT dataset introduced in this work, the above datasets rely on commercial, preprocessed high-resolution imagery, which limits their applicability for real-world Sentinel-2 satellite observations. While they expand the number of classes, they still suffer from a low number of images per class. Their very high resolution (up to 30 cm per pixel) allows distinguishing fine details, such as individual buildings, making them less comparable to the EuroSAT dataset.

A dataset more similar to EuroSAT is the Brazilian Coffee Scene (BCS) dataset, developed by Penatti et al. [8], which uses satellite images with a 10-meter resolution to classify coffee crops. It includes two classes (coffee crop and non-coffee crop), each with 1,423 images containing red, green, and near-infrared bands.

Similarly, Basu et al. [1] [9] introduced the SAT-6 dataset based on aerial imagery from the National Agriculture Imagery Program (NAIP) at 1-meter resolution. SAT-6 includes six classes: barren land, trees, grassland, roads, buildings, and water bodies. The dataset consists of image patches with red, green, blue, and near-infrared bands.

B. Land Use and Land Cover Classification

Convolutional Neural Networks (CNNs) [10] have become the state-of-the-art method for image classification tasks, including LULC classification. Their ability to learn hierarchical features directly from raw pixel data has enabled high accuracy in both RGB and multi-spectral remote sensing datasets [11], [12], [13].

In this study, we implement EfficientNetB0 on the RGB EuroSAT dataset. By using data augmentation and class balancing, our model achieves a final test accuracy of 88.3%, demonstrating that CNN-based architectures can effectively classify medium-resolution satellite imagery with diverse land cover types.

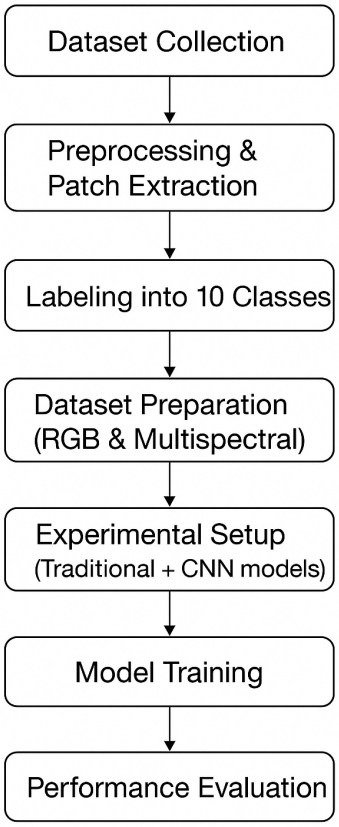

Fig. 3. The diagram illustrates the EuroSAT dataset creation process.

III. Dataset Acquisition

Classifying remotely sensed images requires a reliable labeled dataset. Traditional methods such as Random Forests, Bag-of-Visual-Words (BoVW), and shallow CNNs have been applied on datasets like UCM [9], SAT-6 [14], and BCS [15]. In deep learning contexts, pretrained CNNs from ImageNet [12] have been fine-tuned to remote sensing tasks, achieving

superior performance compared to classical methods [16], [17], [18].

In this study, we leverage the RGB EuroSAT dataset https://www.kaggle.com/datasets/pranjallk1995/rgbeurosat, a collection of medium-resolution satellite images sourced from Sentinel-2, focused on 10 land use and land cover classes. Unlike the full multispectral Sentinel-2 imagery, our work uses only the RGB channels.

To construct the EuroSAT dataset, we performed two main steps:

A. Dataset Preparation

The EuroSAT dataset contains 10 classes: Industrial, SeaLake, Residential, HerbaceousVegetation, River, Pasture, Forest, AnnualCrop, Highway and PermanentCrop. Each class originally included 2,000–3,000 images. To ensure balanced representation for training deep learning models, we applied Random Over-Sampling to equalize the number of images per class at 2,400 images, resulting in a balanced dataset suitable for robust model training.

The dataset was then split into train, validation, and test sets:

- Training set: 3,000 images (300 per class)
- Validation set: 1,000 images (100 per class)
- Test set: 1,000 images (100 per class)

Images were preprocessed using the EfficientNetB0 preprocessing pipeline, including resizing to 64x64 pixels, normalization, and data augmentation techniques such as rotation, horizontal flipping, shifting, and zooming.

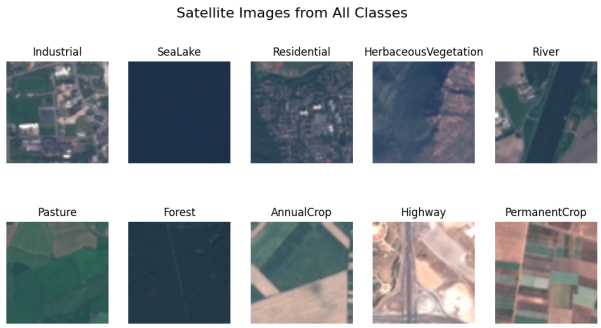

Fig. 4. This overview shows sample image patches of all 10 classes covered in the proposed EuroSAT dataset.The images measure 64x64 pixels. Each class contains 2,000 to 3,000 image. In total, the dataset has 27,000 geo-referenced images.

B. Dataset Characteristics

The dataset provides RGB satellite images in the visible spectrum, with sufficient intra-class variability to support generalizable model training. Each image is geo-referenced, and classes were selected based on both visibility in RGB imagery and relevance to common land use categories. The dataset includes:

- Agricultural land: Annual crops, Permanent crops, Pastures
- Built-up areas: Residential buildings, Industrial buildings, Highways
- Natural regions: Forest, Herbaceous vegetation
- Water bodies: Rivers, Sea/Lakes

Through this preprocessing and sampling, the dataset became well-suited for training deep CNNs, such as EfficientNetB0, achieving strong classification performance on the RGB EuroSAT dataset while maintaining a balanced representation of all 10 classes.

Band	Description	Spatial Resolution	Central Wavelength
B02	Blue	10	490
B03	Green	10	560
B04	Red	10	665

TABLE I
OVERVIEW OF THE RGB BANDS USED FROM THE EUROSAT DATASET FOR EFFICIENTNETB0 CLASSIFICATION.

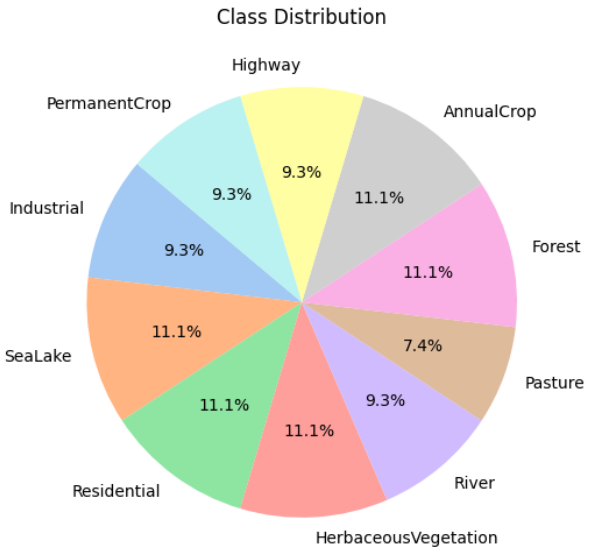

Fig. 5. Class distribution of the EuroSAT Dataset"

IV. DATASET BENCHMARKING

Previous studies have shown that deep CNNs outperform non-deep learning approaches in land use and land cover (LULC) image classification. In our work, we benchmarked the RGB EuroSAT dataset using EfficientNetB0, a state-of-the-art CNN architecture.

The dataset was balanced across 10 classes and split into training (3,000 images), validation (1,000 images), and test (1,000 images) sets. Images were resized to 224×224, normalized using preprocess input, and augmented with rotations, flips, width/height shifts, and zooms to improve generalization.

In our experiments, the EfficientNetB0 model demonstrated promising performance. It achieved a training-validation accuracy of 86.6%, while the test accuracy reached 88.0%. This indicates that the model generalizes well to unseen data and maintains consistency between training and evaluation phases.

A. Model Training and Performance

EfficientNetB0 was fine-tuned with pre-trained ImageNet weights, freezing the majority of base layers while allowing the last few layers to adapt to the EuroSAT dataset. Fully connected layers were added on top, along with batch normalization and dropout layers to prevent overfitting. The model was trained using a sparse categorical cross-entropy loss function and the Adam optimizer, with early stopping and model checkpointing to retain the best validation accuracy.

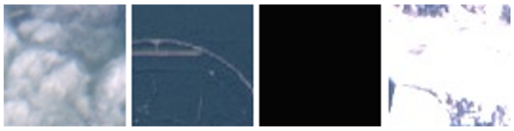

Fig. 6. Four examples of bad image patches, which are intended to show industrial buildings. Clearly, no industrial building is shown due to clouds, mislabeling, dead pixels or ice/snow.

B. Evaluation

The model achieved a final test accuracy of 88.3%, with training accuracy reaching 86.2% and validation accuracy 86.6%. Classification reports indicate strong precision, recall, and F1-scores across most classes, though some confusion occurred between visually similar classes such as AnnualCrop vs Pasture and Forest vs HerbaceousVegetation. The confusion matrix highlights that the model reliably predicts major land-use types while minor misclassifications are limited.

	Precision	Recall	F1-Score
Industrial	0.95	0.98	0.00
SeaLake	0.90	0.88	0.00
Residential	0.90	0.87	0.00
HerbaceousVegeta-	0.95	0.96	0.00
River	0.90	0.93	0.00
Pasture	0.97	0.70	0.00
AnnualCrop	0.90	0.88	0.00
Highway	0.90	0.88	0.00
PermanentCrop	0.98	0.88	0.00
macro avg	0.90	0.88	0.88
weighted avg	0.90	0.88	0.88

Fig. 7. Classification report.

These results demonstrate that EfficientNetB0, even on a subset of the EuroSAT dataset with balanced classes and extensive augmentation, provides robust classification capabilities for satellite imagery.

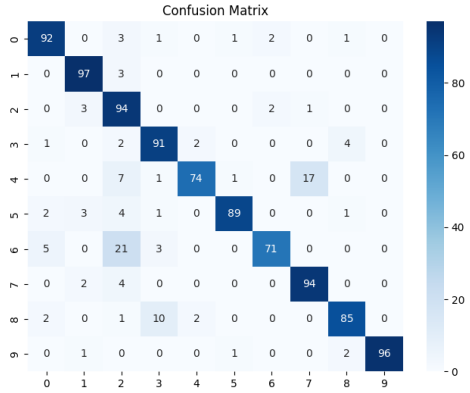

Fig. 8. Confusion matrix of a fine-tuned EfficientNetB0 on the proposed EuroSAT dataset using satellite images in the RGB color space.

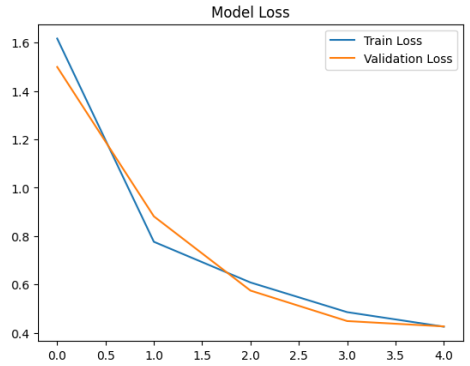

Fig. 9. Representing the history of training and validation loss accross the epochs

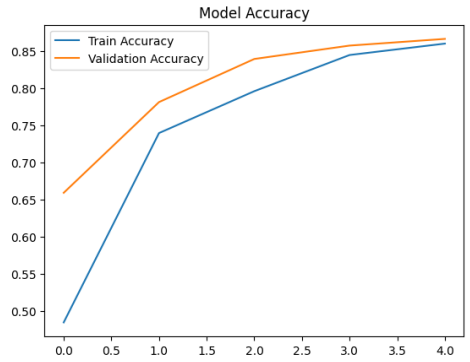

Fig. 10. Representing the history of training and validation accuracy across the epochs

V. APPLICATIONS

The trained EfficientNetB0 model has several practical applications in remote sensing, environmental monitoring, and geospatial analysis:

A. Land Use and Land Cover Change Detection

By analyzing satellite images captured over time, the model can identify changes in land cover. Image patches from the

same geographic location taken at different timestamps are compared, and any difference in predicted labels indicates a change in land usage. This is useful for:

Urban monitoring: Tracking the expansion of residential or industrial areas.

Environmental conservation: Detecting deforestation, wetland degradation, or changes in pasture and forest areas.

Agricultural monitoring: Observing crop rotation, seasonal growth patterns, and changes in farmland.

- Shanghai, China: An area classified as industrial in December 2015 was demolished by December 2016 (Fig. 11).
- Dallas, USA: A region with no dominant residential buildings in August 2015 was developed with new residential buildings by March 2017 (Fig. 12).
- Villamontes, Bolivia: Deforestation was detected between October 2015 and September 2016 (Fig. 13).

These examples are valuable for monitoring urban development, nature conservation, and sustainable management. Detecting deforestation, for instance, is critical for addressing climate change and illegal forest clearing.

Fig. 11. The left image was acquired in the surroundings of Shanghai in December 2015 showing an area classified as industrial. The right image shows the same region in December 2016 revealing that the industrial buildings have been demolished

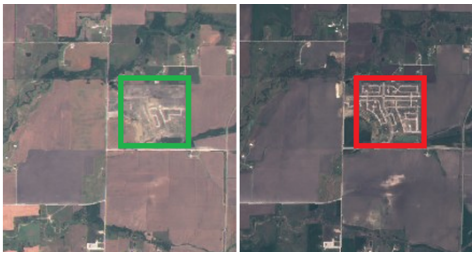

Fig. 12. The left image was acquired in the surroundings of Dallas, USA in August 2015 showing no dominant residential buildings in the highlighted area. The right image shows the same area in March 2017 showing that residential buildings have been built up.

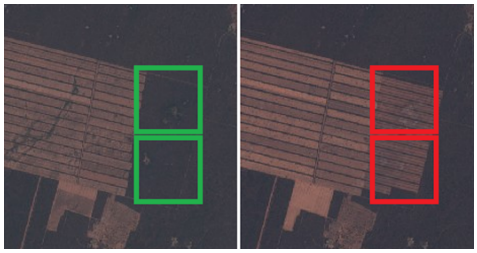

Fig. 13. The left image was acquired near Villamontes, Bolivia in October 2015. The right image shows the same area in September 2016 revealing that a large land area has been deforested.

B. Assistance in Mapping

Although patch-based classification does not provide per-pixel segmentation, it can support keeping maps up-to-date. Systems using the trained CNN can verify existing map data, identify mislabeling, or enable large-scale tagging, such as for crowdsourced maps like OpenStreetMap (OSM).

For example, Fig. 14 shows industrial areas in two regions. In Australia, the OSM map closely matches the satellite image, whereas in Shanghai, a significant portion of industrial areas is not yet mapped. With the frequent revisit of Sentinel-2 imagery, the proposed system combined with EuroSAT can assist in updating and maintaining accurate maps. Further analysis can be achieved using high-resolution imagery and advanced segmentation methods [19], [12].

Fig. 14. A patch-based classification system can verify already tagged areas, identify mistagged areas or bring large area tagging as shown in the above images and maps. The left Sentinel-2 satellite image was acquired in Australia in March 2017. The right satellite image was acquired in the surroundings of Shanghai, China in March 2017. The corresponding up-to-date OpenStreetMap (OSM) mapping images show that the industrial areas in the left satellite image are almost completely covered (colored gray). However, the industrial areas in the right satellite image are not properly covered.

Frequent revisits of Sentinel-2 satellites allow continuous monitoring, enabling up-to-date and reliable mapping. This can be particularly important in regions undergoing rapid urbanization or environmental changes.

VI. CONCLUSION

This study applied EfficientNetB0 to the RGB EuroSAT dataset to classify 10 distinct land-use and land-cover classes. The model achieved 88.3% test accuracy, demonstrating strong capability for identifying diverse land types, even on a moderately sized and balanced dataset.

While this accuracy is lower than some benchmarks reported in the literature (98%), it reflects the performance on a smaller, well-controlled dataset with augmentation.

The trained model offers practical applications in:

Monitoring environmental and urban changes over time.

Supporting mapping and verification tasks in geographical information systems.

Providing a foundation for future work involving higher-resolution satellite imagery, multi-spectral analysis, or pixel-level segmentation models.

Overall, this work demonstrates that EfficientNetB0 is a reliable tool for satellite image classification and lays the groundwork for automated, large-scale land monitoring and geospatial analysis.

REFERENCES

[1] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pp. 270–279, 2010.

[2] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *International journal of remote sensing*, vol. 33, no. 8, pp. 2395–2412, 2012.

[3] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[4] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître, "Structural high-resolution satellite image indexing," in *ISPRS TC VII Symposium-100 Years ISPRS*, vol. 38, pp. 298–303, 2010.

[5] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *Journal of Applied Remote Sensing*, vol. 10, no. 3, pp. 035004–035004, 2016.

[6] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 197–209, 2018.

[7] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[8] O. A. Penatti, K. Nogueira, and J. A. Dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 44–51, 2015.

[9] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "Deepsat: a learning framework for satellite imagery," in *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, pp. 1–10, 2015.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, pp. 630–645, Springer, 2016.

[12] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 1–9, 2016.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[14] C. De Boor and C. De Boor, *A practical guide to splines*, vol. 27. springer New York, 1978.

[15] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[16] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[17] K. Ni, R. Pearce, K. Boakye, B. Van Essen, D. Borth, B. Chen, and E. Wang, "Large-scale deep learning on the yfcc100m dataset," *arXiv preprint arXiv:1502.03409*, 2015.

[18] K. Nogueira, O. A. Penatti, and J. A. Dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539–556, 2017.

[19] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel, "Multi-task learning for segmentation of building footprints with deep neural networks," in *2019 IEEE international conference on image processing (ICIP)*, pp. 1480–1484, IEEE, 2019.