HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 4 (2023)

# Improving Deforestation Detection Accuracy in Noisy Satellite Images with Contrastive Learning-based Approach

Jitae Kim[1], Lim Lee[1], Sihu Park[2#] and Lenny Musunga[1#]

[1]Korea International School – Pangyo Campus, South Korea
[2]North London Collegiate School Jeju, South Korea
[#]Advisor

## ABSTRACT

Deforestation, the large-scale destruction of trees, has far-reaching biological and environmental consequences that pose a significant threat to the environment. Accurate deforestation detection is crucial for successful conservation initiatives and effective land management. Over the last decade, numerous deforestation detection methods utilizing spaceborne photography have been proposed. However, these methods tend to be sensitive to unique image noise in the satellite domain by virtue of the diverse aerial characteristics and air qualities in different regions. To solve this problem, we propose a novel noise-robust deforestation detection framework with a contrastive-learning based approach. The proposed framework consists of two phases: contrastive learning, which aims to extract similar feature embeddings for the same category, proceeded with transfer learning in order to develop the deforestation classifier. Remarkably, the proposed contrastive learning approach successfully handles noisy input satellite images during the feature extraction process. Upon conducting validation, we have found that the proposed method outperforms existing deforestation detection methods by a significant performance gap, highlighting the effectiveness of the proposed contrastive learning approach.

## 1. Introduction

### 1.1 Problem Definition

Deforestation is a major environmental problem that affects the world in numerous detrimental ways. The loss of forest cover causes the degradation of natural habitats, contributing to the decline of biodiversity, water quality, soil fertility, and climate stability. Furthermore, deforestation is also responsible for a significant amount of greenhouse gas emissions that contribute to global warming.

In recent years, the use of satellite imagery has become a popular method for detecting and monitoring deforestation. By analyzing satellite images, scientists are able to identify changes in forest cover over time and determine the extent of deforestation in different regions. This information can then be used to help policymakers develop effective strategies to combat deforestation and promote sustainable land use practices.

The use of satellite imagery for deforestation detection offers several distinct advantages over traditional ground-based methods. Satellite imagery can cover large areas swiftly, enabling large expanses of forests and landscapes to be observed. Additionally, satellite imagery can also be used to detect changes in forest cover in remote areas that are difficult to access by ground-based methods.

However, detecting deforestation in satellite imagery can be a challenging task. It requires the use of advanced image processing techniques to distinguish between natural changes in vegetation cover and human-induced deforestation. Furthermore, the accuracy of deforestation detection depends on several factors, such as the quality of the satellite images, the availability of ground-based data for validation, and the complexity of the forest landscape.

### 1.2 Previous Method

To address the challenge of accurate deforestation detection using satellite imagery, numerous research studies have proposed approaches centered around convolutional neural networks (CNNs). Ortega et al. for instance, demonstrated the feasibility of exploiting convolutional neural networks to develop a deforestation detection system (Ortega et al. 2019). To enhance the accuracy of the deforestation detection system, Ortega et al. proposed a unified machine learning framework that involves the fusion of both traditional machine learning methods and deep learning methods (Ortega et al. 2020). Additionally, John et al. presented a UNet-based (Ronneberger et al. 2015) approach for detecting deforestation areas with pixel-level accuracy (John et al. 2022). However, these methods tend to produce inaccurate results in external validation due to their heavy bias towards the training dataset and sensitivity to image noise commonly present in satellite images.

### 1.3 Proposed Method

In this research, we propose a novel contrastive learning-based deforestation detection framework to effectively leverage unlabeled data and improve the accuracy and robustness of the deforestation detection system. The proposed framework is composed of two phases: contrastive learning and transfer learning. In the contrastive learning phase, the main goal is to train a convolutional neural network to extract consistent feature vectors. By doing so, we address the issues of noise suppression in satellite images and improve the model's generalization capacities. Subsequently, this well-generalized convolutional neural network serves as a “good starting point” for training the deforestation detection model in the transfer learning phase. Through experiments, we have found that employing the proposed transfer learning approach to train a deforestation detection model yields better results compared to training the CNN from scratch.

The following chapters are structured as follows: Chapter 2 introduces the background knowledge to facilitate a better understanding of the proposed method, Chapter 3 provides a comprehensive overview on the development of the proposed method, Chapter 4 demonstrates the superiority of the proposed model through extensive experimentation, and Chapter 5, imparts a summary of the entire research paper.

## 2. Related Work

### 2.1 Image Classification

Image classification is a branch of computer vision and machine learning which extracts the features of images and assigns it to one of several predefined classes. At the core of image classification systems, convolutional neural networks (CNNs) play a pivotal role due to their ability to extract various characteristics from inputted images.

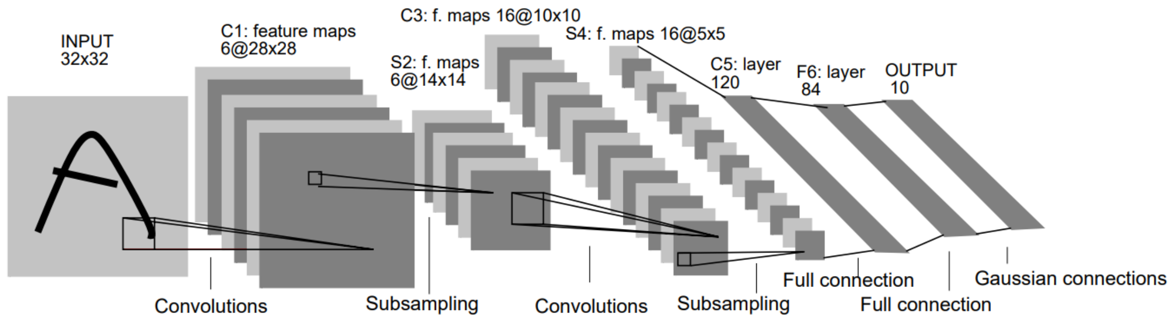



**Figure 1.** Image Classification Example (LeCun et al. 1998)

Figure 1 illustrates an example of an image classification system developed using a convolutional neural network. As shown, the network takes digital images as input and extracts feature maps containing essential visual patterns and information to accurately assign the input to a predefined category.

Similarly, in this research, we consider deforestation detection as an image classification task, employing a system that inputs a batch of satellite images and classifies them into one of two categories, either "deforestation present" or "deforestation not present". The proposed module utilizes image classification to extract the features of a forested image of earth, taking the forest-to-land ratio into consideration, and subsequently uses this data to train the deforestation-detection module using a process deemed "contrastive learning". The detailed information on how we develop the proposed method will be further explained in Chapter 3.

## 2.2 Deforestation Detection

Deforestation is the process of cutting down a wide area of trees, enough to be identified in satellite imagery. Deforestation Detection is the process of identifying the presence of deforestation in satellite images. The consequences of deforestation are far-reaching, impacting agriculture, ecology, and the environment of the surrounding area. There have been numerous studies utilizing machine learning algorithms to detect deforestation, such as CNN or supervised learning methods.

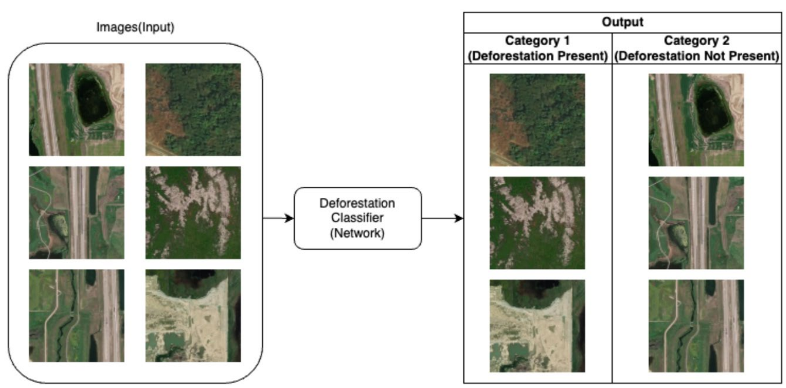



**Figure 2.** Deforestation Detection Example

As shown in fig. 2, the following paper proposes an accurate deforestation detection system which classifies an inputted satellite image into two categories (either deforestation present or not present) without any bias. This process will be elaborated on further in chapter 3.

# 3. Approach

In this chapter, we provide a detailed review of the proposed deforestation detection framework, consisting of two main phases: contrastive learning and transfer learning. In contrastive learning, a convolutional neural network is trained to extract consistent feature vectors for noise suppression and improved generalization. During training, the network is penalized based on the similarity score between images belonging to the same category, whether they depict deforestation or not. In transfer learning, the pretrained convolutional neural network is fine-tuned to accurately classify the input satellite image. Leveraging the pretrained network's ability to extract consistent features, the trained deforestation classifier then produces robust results.

## 3.1 Phase 1: Contrastive Learning

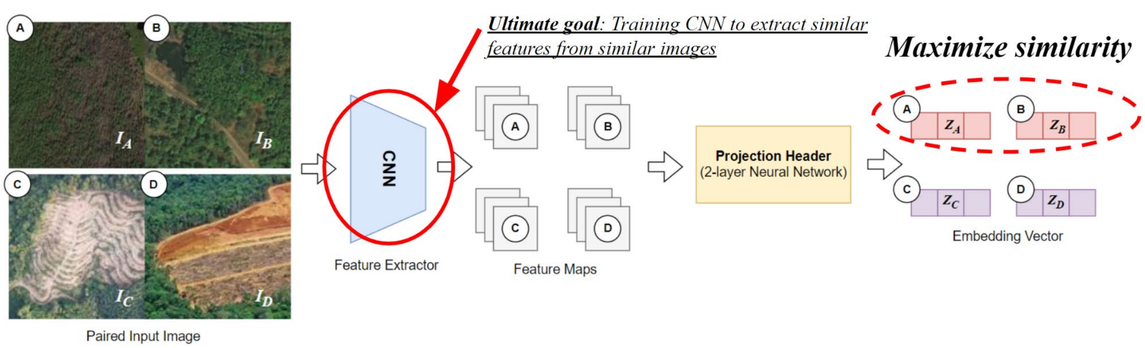



**Figure 3.** Overall Architecture of the Proposed Methods (phase1)

The proposed framework takes multiple images from a given set of image samples, denoted as $\boldsymbol{I} \in \boldsymbol{R}^{HW}$ , where H and W represent the height and width of the input satellite image $\boldsymbol{I}$. These image pairs are then fed into a convolutional neural network to extract feature maps. More specifically, the image pairs first pass through a feature extractor, ***FeatExt***, which is developed using a convolutional neural network, to extract feature map, ***feat***. We define the proposed Feature Extractor as : ***FeatExt: I → feat***. After this, the extracted feature maps, ***feat,*** are inputted to the projection header to create the embedding vectors $\boldsymbol{Z}$. To determine the similarity between the embedding vectors, a cosine similarity function is utilized. If images belong to the same category, their corresponding embedding vectors should exhibit high similarity.

**Equation 1**: Cosine Similarity Function

$$S_{A,B} = \frac{{Z_A}^T Z_B}{(\tau \|Z_A\| \|Z_B\|)}$$

As seen in Equation 1, $\boldsymbol{S_{A,B}}$ denotes the similarity score between two embedding vectors $\boldsymbol{Z_A}$ and $\boldsymbol{Z_B}$. These scores indicate how similar the feature vectors are – higher scores suggest images belong to the same disease category and share similar features. Indeed, as shown, a similarity score shows, at a glance, the correspondence of two selected feature vectors. After calculating all the similarities between embedding vectors, the loss value is computed using the log-sum-exponential function. The loss function takes the logarithm of the sum of the exponential of the similarity scores for all pairs. This function encourages similar images to have higher scores while penalizing dissimilarity, contributing to accurate deforestation detection.

**Equation 2**: Contrastive Learning Loss Function

$$L = -\log\left(\frac{e^{S_{A,B}}}{e^{S_{A,B}} + e^{S_{A,C}} + e^{S_{A,D}}}\right)$$

Here, $S_{i,j}$ denotes the pre-calculated similarity score between embedding $Z_i$ and $Z_j$. The loss function is like the cross-entropy loss used in classification models. As the similarity score between the embeddings extracted from the same category is maximized, the cross-entropy loss function can subsequently be employed to train the network.

## 3.2 Phase 2: Transfer Learning


**Figure 4.** Overall Architecture of the Proposed Methods (phase2)

Figure 4 demonstrates the second phase of the proposed deforestation detection framework. In this phase, the transfer learning approach is utilized to train the deforestation classifier, aiming to determine whether the input image belongs to the deforestation category or not. The pretrained Feature Extractor takes the input image $I \in R^{HW}$, and produces feature map ***feat***. Here, we should note that this feature extraction process is identical to the feature extractor used in phase 1. The feature map ***feat*** is then fed to the Deforestation Classifier to predict the category score. This score determines the presence of deforestation by converting the predicted category score to probability and subsequently calculating the loss value using Equation 3.

**Equation 3**: Cross-Entropy Loss Function

$$L = -\log_e P$$

Here, $P$ denotes the predicted probability, which plays a crucial role in the proposed transfer learning approach. This process enables effective training on the deforestation classification task, resulting in significantly improved performance. A detailed explanation of the effectiveness of the proposed approach will be provided in Chapter 4.

# 4. Results

This chapter presents a comprehensive account of the experimental process, including details about the training and testing dataset used, evaluation metrics, and the procedure of the experiments conducted.

## 4.1 Dataset

In this study, we utilize two publicly available deforestation satellite datasets: one acquired in Brazil (Torres et al., 2021) and the other collected in Ukraine (Isaienkov et al., 2021). Both datasets contain labeled samples categorized as either representing deforestation present or not. There are 61,021 positive (deforestation) samples and 57,768 negative samples. We divided the samples into an 80:20 ratio, with 80% allocated for training and 20% for testing.

## 4.2 Experimental Protocol

To assess the effectiveness of the proposed methods, we conduct two experiments: comparisons with state-of-the-art methods, and evaluations utilizing data augmentation. To ensure a comprehensive evaluation of the proposed module, we employed four performance metrics: accuracy, recall, precision, and F1 score. The equations for these metrics are shown in Equations 4-7.

**Equation 4**: Accuracy

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Samples}$$

**Equation 5**: Recall

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

**Equation 6**: Precision

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

**Equation 7**: F1-Score

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Here, the true positive denotes the instances where the proposed module correctly predicts the "positive" class, which are deforestation-present images. In contrast, the true negative variables show the instances where the proposed module correctly predicts the "negative" class, where images are non-deforested. Vice versa, false positives and false negatives are the occurrences of when the module incorrectly predicts the "positive" and "negative" classes.

## 4.3 Comparison with state-of-the-art method

**Table 1**. Comparison with state-of-the-art deforestation methods.

| Method | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| AlexNet (Krizhevsky et al. 2012) | 0.6761 ± 0.0011 | 0.7244 ± 0.0009 | 0.6730 ± 0.0008 | 0.6978 ± 0.0009 |
| VGG19 (Szegedy et al. 2014) | 0.7731 ± 0.0006 | 0.7943 ± 0.0011 | 0.7984 ± 0.0008 | 0.7963 ± 0.0011 |
| Resnet18 (He et al. 2016) | 0.7803 ± 0.0005 | 0.7918 ± 0.0009 | 0.8201 ± 0.0007 | 0.8057 ± 0.0010 |
| HRNet32 (Cheng et al. 2020) | 0.8151 ± 0.0003 | 0.8375 ± 0.0006 | 0.8303 ± 0.0008 | 0.8830 ± 0.0009 |
| Xception based (Torres et al. 2021) | 0.8274 ± 0.0004 | 0.8476 ± 0.0011 | 0.8404 ± 0.0009 | 0.8830 ± 0.0009 |
| Proposed Method | **0.8771 ± 0.0008** | **0.8820 ± 0.0012** | **0.8990 ± 0.0014** | **0.8904 ± 0.0009** |

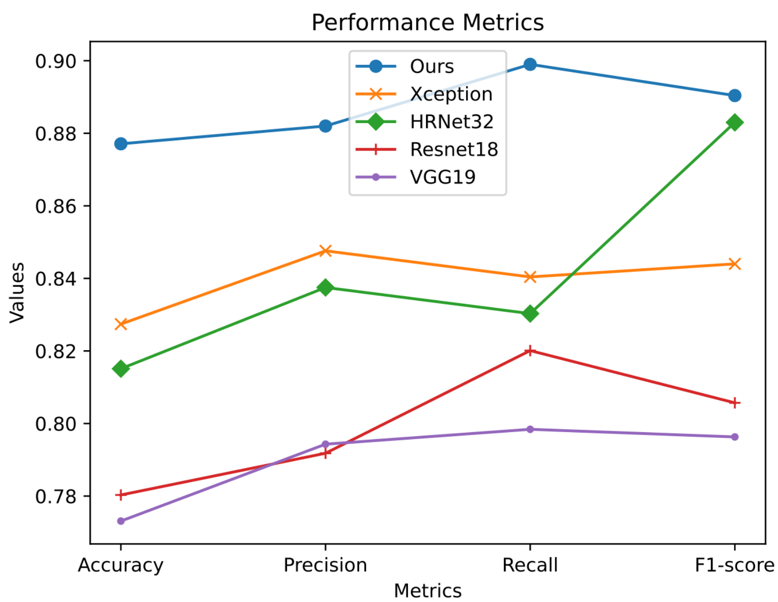
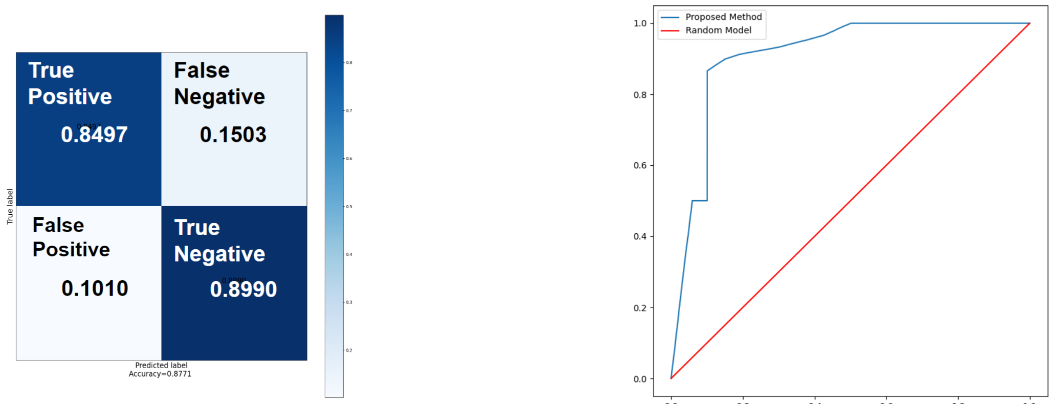



**Figure 5.** Visual comparison of performance metric with state-of-the-art deforestation methods.

Table 1 demonstrates the performance comparison between the proposed method and the selected state-of-the-art methods. We chose AlexNet (Krizhevsky et al. 2012), VGG (Szegedy et al. 2014), Resnet18 (He et al. 2016), HRNet32 (Cheng et al. 2020), and Xception (Torres et al. 2021) for the comparison methods –, these classifiers all show comparable performance in image classification tasks.

As shown in Table 1, the proposed method outperforms the state-of-the-art methods in every evaluation metric. We attribute this success to the proposed contrastive learning approach – compared to the proposed module, the comparison methods tend to exhibit bias towards the training dataset and yield poor results for testing because they are trained in a supervised manner.

When analyzing the accuracy of a convolutional neural network, it becomes evident that increased depth leads to an improved performance. Indeed, this pattern is observed between shallower networks like ResNet18 and VGG, and deeper networks, such as Xception and HRNet32 as well. Despite the proposed method having a similar network depth to the group of deeper networks, the proposed network shows superior performance compared to the deeper networks trained using a supervised approach. This finding clearly shows and proves the effectiveness of the proposed contrastive learning approach.




(a) (b)

**Figure 6.** Confusion matrix and Receiver Operating Characteristic (ROC) curve of the proposed method (a): Confusion matrix of the proposed methods, and (b): ROC curve.

Figure 6 demonstrates the confusion matrix of the proposed method. The value for true positive is 0.8497, false negative is 0.1503, False positive is 0.1010, and true negative is 0.8990. As the diagonal component of the confusion matrix demonstrates a notably higher ratio, signifying a substantial number of correct predictions, it is proven that the proposed method shows its robustness in handling satellite images with noise.

## 4.4 Ablation Study

**Table 2**. Ablation Study Result

| Method | Accuracy |
|---|---|
| w/o data augmentation | 0.8421 |
| Grayscale | 0.8141 (-2.8%) |
| Gaussian Blur | 0.8041 (-3.8%) |
| Horizontal Flip | 0.8510 (+0.9%) |
| Gaussian Noise | 0.8531 (+1.1%) |
| Color Jitter | 0.8731 (+3.1%) |
| Horizontal Flip + Gaussian Noise + Color Jitter | **0.8771 (+3.5%)** |

Table 2 shows the results of the ablation study conducted on each data augmentation technique utilized in this research. The purpose of this experiment is to gain insight into the importance of each data augmentation method for overall performance. As a baseline, we initially train the proposed network without using any data augmentation techniques. Subsequently, we investigate five data augmentation techniques often used to increase the variety of the training samples in image classification tasks.

The grayscale conversion technique resulted in an accuracy of 0.8141, a decrease of 2.8% compared to the baseline. This is due to grayscale conversion reducing the color information in the images, leading to a loss of discriminative features and confusing the module.

The gaussian blur technique yielded an accuracy of 0.8041, indicating a decrease of 3.8% compared to the baseline. Gaussian blur blurs important details and makes it harder for the proposed module to accurately classify them.

However, flipping the images horizontally increased the accuracy to 0.8510, with a 0.9% improvement over the baseline. Horizontal flipping, in this case, provided additional training examples with reversed orientations, and helped the module learn to be more robust to image orientation variations.

Applying gaussian noise to the images resulted in an accuracy of 0.8531, with a 1.1% improvement compared to the baseline. Gaussian noise added random variations to the pixel values, and enhanced the module's ability to generalize and handle noisy input data.

Additionally, the color jitter augmentation led to an accuracy of 0.8731, with a 3.1% improvement over the baseline. Color jitter introduces random color transformations, which can help the system become more robust to variations in color distribution. In the end, the proposed module showed a 3.5% improvement compared to the baseline and resulted in an accuracy of 0.8771 by combining horizontal flipping, gaussian noise, and color jitter.

## 5. Conclusion

In this research paper, we investigated the utilization of contrastive learning to develop a more robust satellite image deforestation-detection system that effectively addresses image noise. We proposed a deforestation detection framework comparing pairs of images together to produce a more accurate classification. Through experimentation, we proved that the proposed module demonstrated superior performance compared to conventional supervised learning methods. The findings of this study have significant implications for future research on techniques that classify satellite images proposing a solution for image noise. We expect the proposed module will find applications in diverse domains, such as image noise reduction and object detection, leading to more precise and unbiased classifications across various fields.

## References

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Isaienkov, K., Yushchuk, M., Khramtsov, V., & Seliverstov, O. (2020). Deep learning for regular change detection in Ukrainian forest ecosystem with sentinel-2. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 364-376. https://doi.org/10.1109/JSTARS.2020.3034186

John, D., & Zhang, C. (2022). An attention-based U-Net for detecting deforestation within satellite sensor imagery. International Journal of Applied Earth Observation and Geoinformation, 107, 102685. https://doi.org/10.1016/j.jag.2022.102685

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Ortega Adarme, M., Queiroz Feitosa, R., Nigri Happ, P., Aparecido De Almeida, C., & Rodrigues Gomes, A. (2020). Evaluation of deep learning techniques for deforestation detection in the Brazilian Amazon and cerrado biomes from remote sensing imagery. Remote Sensing, 12(6), 910. https://doi.org/10.3390/rs12060910.

Ortega, M. X., Bermudez, J. D., Happ, P. N., Gomes, A., & Feitosa, R. Q. (2019). Evaluation of deep learning techniques for deforestation detection in the amazon forest. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4, 121-128

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9). https://doi.org/10.48550/arXiv.1409.4842

Torres, D. L., Turnes, J. N., Soto Vega, P. J., Feitosa, R. Q., Silva, D. E., Marcato Junior, J., & Almeida, C. (2021). Deforestation detection with fully convolutional networks in the Amazon forest from Landsat-8 and Sentinel-2 images. Remote Sensing, 13(24), 5084. https://doi.org/10.3390/rs13245084

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3349-3364. https://doi.org/10.1109/TPAMI.2020.2983686