A Project report on

Uncertainty in Deep Learning Models for Building Detection In Satellite Images

Submitted by

Mr Pranav Manohar Pathekar

BAJAJ INSTITUTE OF TECHNOLOGY, WARDHA

under the esteemed guidance of

Miss. Reedhi Shukla & Mr. Sampath Kumar

Scientist/Engineer

NATIONAL REMOTE SENSING CENTRE
INDIAN SPACE RESEARCH ORGANISATION
DEPT OF SPACE, GOV OF INDIA
BALANAGAR, TELANGANA
(2024)

भारत सरकार
अन्तरिक्ष विभाग
राष्ट्रीय सुदूर संवेदन केन्द्र
बालानगर, हैदराबाद-500 037, तेलंगाना, भारत
टेलिफोन : +91 40 23878360
+91 40 23884000-04
फैक्स : +91 40 23877210

Government of India
Department of Space
National Remote Sensing Centre
Balanagar, Hyderabad-500 037, Telangana, India
Telephone : +91 40 23878360
+91 40 23884000-04
Fax : +91 40 23877210

Certificate

This is to certify that this project report entitled **"Uncertainty in Deep Learning Models for Building Detection In Satellite Images"** is a record of the bonafide work carried out by Mr Pranav Manohar Pathekar under our guidance and supervision at National Remote Sensing Centre, ISRO, Hyderabad during the time period march 2024 to May 2024. This project report is beng submitted in partial fulfilment of B.Tech in CSE at Bajaj Institute Of Technology, Wardha. He has completed the assigned work satisfactory.

Miss.Reedhi Shukla	**Mr. Sampath Kumar**
Scientist/Eng,'SE'	Scientist/Eng,'SF'

Declaration

I hereby declare that the project titled "Uncertainty in Deep Learning Models for Building Detection In Satellite Images" is carried out by me under the guidance of **Miss. Reedhi Shukla**,Sci/Engr,'SE' and **Mr.Sampath Kumar**,Sci/Engr,'SF'. There results embodied in this project work have not been submitted by any another university or institute for the award of any degree.

Report Title: Uncertainty in Deep Learning Models for Building Detection In Satellite Images

SN	**Student Name**	**Signature**
1	Pranav Manohar Pathekar	

Date: May 30, 2024
Place: NRSC Jeedimetla

Acknowledgment

- I would like to express my gratitude and appreciation to all those who gave me the possibility to complete this report. Special thanks to my project guide **Miss. Reedhi Shukla**,Sci/Engr,'SE' and **Mr.Sampath Kumar**,Sci/Engr,'SF' who helped , stimulating suggestions and encouraged me all time of process and in completing my project.

- My deepest thanks to **Prof. Sheetal Kale**, Head of the Department of Computer Engineering and **Dr. Narendra Kanhe**, Principal Bajaj of Institute of Technology, Wardha.

- I express our sincere thanks to **Dr.PRAKASH CHAUHAN**, Director of NRSC for giving me an opportunity and providing all necessary support in carrying out my internship at NRSC.

- I would also like to acknowledge with much appreciation the crucial role of the staff of NRSC.

- Finally, I would like to extend my deep sense of gratitude to all my friends, and last but not least I am greatly indebted to my parents who inspired me under all cricumstances.

SN	**Student Name**	**Signature**
1	Pranav Manohar Pathekar	

Abstract

- Satellite imagery provides a valuable data source for monitoring urban environments and detecting buildings. However, applying deep learning models to satellite imagery for building detection remains challenging due to inherent uncertainties in the data and models. This study investigates the uncertainties associated with deep learning models for building detection in satellite images, with a focus on the U-Net architecture.

- We explore the impact of various hyperparameters on model performance and assess the model's confidence through extensive tuning and evaluation. By systematically varying hyperparameters such as learning rate, batch size, and loss functions, we quantify their effect on the model's accuracy and uncertainty estimates. To facilitate comprehensive analysis and visualization of results, we leverage WandB platform for experiment tracking and monitoring. Our findings provide insights into the robustness and reliability of deep learning models for building detection tasks, informing strategies for mitigating uncertainties, and improving model performance. Additionally, this study contributes to the broader understanding of uncertainty quantification in deep learning applications for remote sensing and computer vision.

Contents

Chapter 1

Introduction

1.1 Background

Building detection in satellite images is a critical task in various fields such as urban planning, disaster management, and navigation systems. The advent of deep learning has significantly enhanced the accuracy and efficiency of building detection tasks. However, the performance of deep learning models can be influenced by uncertainties arising from data quality and model parameters. Understanding and mitigating these uncertainties is crucial for reliable and robust building detection systems.

1.2 Types of Uncertainty in Deep Learning

Uncertainty in deep learning models can be broadly classified into two categories:

- 1. Data Uncertainty: Also known as aleatoric uncertainty, it arises from the inherent noise in the data. This type of uncertainty can be addressed by improving data quality.
- 2. Model Uncertainty: Also known as epistemic uncertainty, it arises from the model parameters and can be reduced through techniques such as hyperparameter tuning.

1.3 Organization of the report

The report is organized as follows:
Chapter 2: Literature Survey
Chapter 3: Methodology
Chapter 4: Implementation
Chapter 5: Results & Discussion
Chapter 6: Conclusion & Future Scope

Chapter 2

Literature Survey

2.1 Literature review

Extensive research has been conducted on building detection in satellite images using various deep learning techniques. This section reviews key studies in the domain to provide a comprehensive understanding of existing methods and identify areas for improvement.

References	**Literature Survey**
International society for photogrammetry and remote sensing.	This source provides comprehensive information on building detection from aerial imagery using inception resnet, unet and unet architectures
Computational Inteligence and Neuroscience.	Detecting Buildings and Nonbuildings from Satellite Images Using U-Net
National Library of Medicine.	Active Learning with Bayesian UNet for Efficient Semantic Image Segmentation
IEEE International Geoscience and Remote Sensing Symposium IGARSS.	Bayesian Deep Learning with Monte Carlo Dropout for Qualification of Semantic Segmentation
International Journal of Performability Engineering,	Hyperparameter Tuning in Deep Learning-Based Image Classification to Improve Accuracy using Adam Optimization

Table 2.1: Summary of Literature Papers

2.2 Gap identification in the literature

Despite the advancements in building detection using deep learning, several gaps remain: Most studies emphasize accuracy improvements without adequately addressing data and model uncertainties. There is limited exploration of the impact of noisy data and the effectiveness of normalization techniques in mitigating data uncertainty. Comprehensive studies integrating uncertainty quantification techniques with model performance optimization are scarce.

2.3 Problem statement

The primary objective of this project is to investigate and mitigate uncertainties in deep learning models for building detection in satellite images. The focus will be on both data uncertainty (arising from noisy data) and model uncertainty (addressed through hyperparameter tuning and Bayesian approaches).

2.4 Challenges

- **Data Format Complexity:** Data format complexity is a significant challenge in many image processing and machine learning projects, including image denoising. This challenge can arise from various factors related to the structure, representation, and variability of the image data.

- **Overfitting:** In the context of image denoising projects using models like U-Net, overfitting can severely impact the model's ability to effectively remove noise from images that were not part of the training set.

2.5 Objectives

The main objectives of the project are as follows:

- To study the impact of model uncertainties on building detection.
- To design experiments for evaluating uncertainties using noisy and non-noisy data.
- To analyze the performance of the U-Net model under different uncertainty conditions.
- To implement hyperparameter tuning techniques to reduce model uncertainty.

2.6 Scope of work

This project involves the following tasks:

- Data preparation and preprocessing.
- Experimentation with non-noisy data.
- Model training and evaluation using the U-Net architecture.
- Analysis of model uncertainties.
- Hyperparameter tuning for uncertainty reduction.

Chapter 3

Methodology

3.1 Data Collection and Preprocessing

The dataset consists of satellite images with corresponding building annotations. The data is divided into noisy and non-noisy sets, and each set is further split into training, validation, and test sets. Both normalized and non-normalized versions of the data are prepared to study the impact of normalization on model performance and uncertainty.the dataset is of satellite images of hyderabad region with its mask

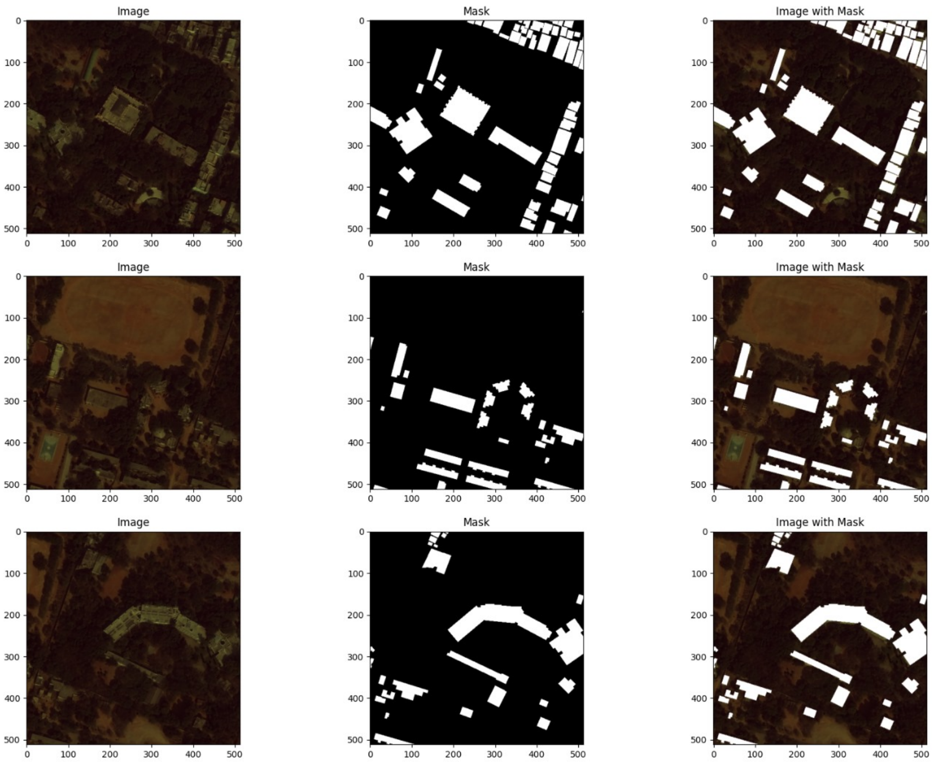

Figure 3.1: Satellite images of hyderabad region with its mask.

3.2 U-Net Architecture

The U-Net architecture is a convolutional neural network designed for biomedical image segmentation. It has a U-shaped structure comprising a contracting path and an expansive path, making it highly effective for image segmentation tasks.The U-Net model, originally designed for biomedical image segmentation, has proven to be highly effective in various image processing tasks, including denoising. Its architecture comprises an encoder-decoder structure with symmetric skip connections, which helps in preserving spatial information and ensuring precise reconstructions. The encoder path captures context through a series of convolutional and pooling layers, while the decoder path reconstructs the image using upsampling and convolutional layers.

3.2.1 Contracting Path

The contracting path consists of a series of convolutional layers, each followed by a rectified linear unit (ReLU) and a max-pooling operation. This path captures the context in the image by progressively reducing the spatial dimensions and increasing the number of feature channels.

3.2.2 Bottleneck

The bottleneck is the bridge between the contracting and expansive paths. It contains convolutional layers with the highest number of feature channels and serves as the core of the network where the most abstract features are captured.

3.2.3 Expansive Path

The expansive path consists of a series of up-convolutional layers that increase the spatial dimensions of the feature maps. It also includes concatenation operations that combine feature maps from the contracting path with the upsampled feature maps to provide localized information for precise segmentation.

3.2.4 Skip Connections

Skip connections between the contracting and expansive paths enable the network to use fine-grained details from earlier layers, improving segmentation accuracy.

3.2.5 Output Layer

The output layer is typically a convolutional layer with a softmax activation function, producing the final segmented image with pixel-wise classification.

[article] graphicx float

3.3 Data Uncertainty Analysis

To analyze data uncertainty, predictions are made on both noisy and non-noisy datasets. Metrics such as accuracy, loss, validation accuracy, and validation loss are recorded. Confusion matrices are generated to quantify the level of data uncertainty.

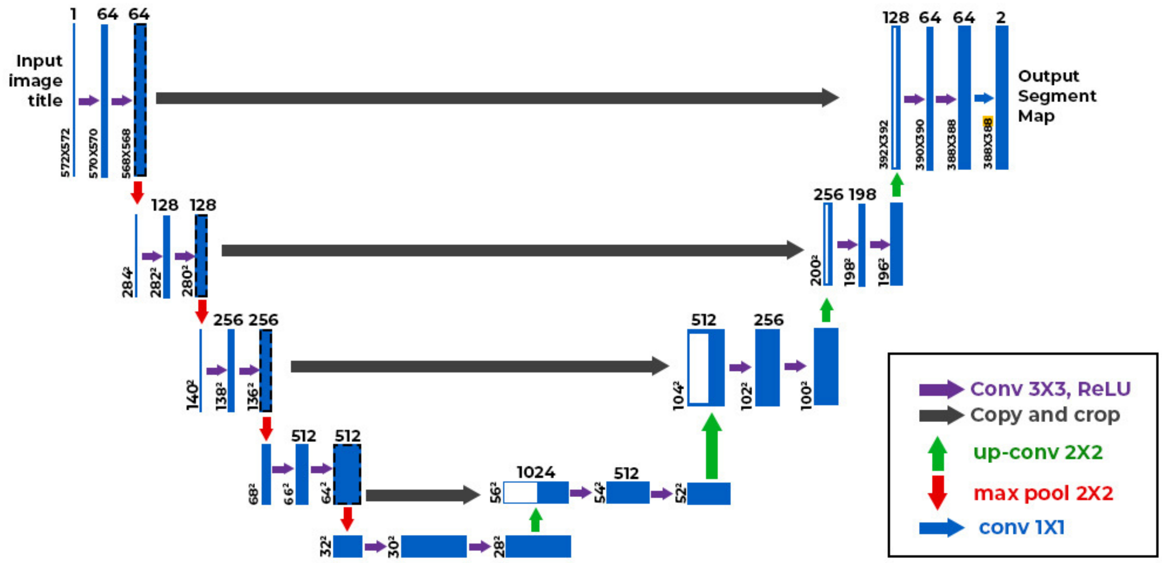

Figure 3.2: U-Net Architecture for Image Segmentation

3.4 Model Uncertainty Analysis

Model uncertainty is addressed by tuning various hyperparameters, including the loss function, activation function, number of hidden layers, and cross-entropy as a measure of uncertainty. Different configurations are tested to identify the optimal settings that reduce uncertainty.

3.5 Evaluation Metrics

The performance of the model is evaluated using the following metrics:

- **Accuracy:** Measures the proportion of correct predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

- **Precision:** Measures the proportion of true positive predictions out of all positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

- **Loss:** Measures the difference between the predicted and actual values.

$$Loss = \frac{1}{N} \sum_{i=1}^{N} L(y_i, \hat{y}_i) \tag{3.3}$$

- **Validation Accuracy:** Measures the model's accuracy on the validation set.

$$ValidationAccuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (onvalidationset) \tag{3.4}$$

- **Validation Loss:** Measures the loss on the validation set.

$$ValidationLoss = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{y}_i) \quad (onvalidationset) \tag{3.5}$$

- **Confidence Scores:** Indicate the model's confidence in its predictions, helping to assess its reliability.

$$ConfidenceScore = \hat{y}_i \quad (predictedprobability)$$

- **Binary Accuracy:** Measures the proportion of correct binary predictions.

$$BinaryAccuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.6}$$

- **Confusion Matrix:** Provides a detailed breakdown of the model's performance, highlighting true positives, false positives, true negatives, and false negatives.
- **False Positives:** Counts the number of negative instances incorrectly classified as positive.
- **False Negatives:** Counts the number of positive instances incorrectly classified as negative.
- **True Positives:** Counts the number of positive instances correctly classified as positive.
- **True Negatives:** Counts the number of negative instances correctly classified as negative.
- **Recall:** Measures the proportion of true positive predictions out of all actual positives.

$$Recall = \frac{TP}{TP + FN} \tag{3.7}$$

- **AUC (Area Under the Curve):** Measures the model's ability to distinguish between classes.

$$AUC = \int_0^1 TPR(FPR)\, dFPR \tag{3.8}$$

Chapter 4

Implementation

4.1 Experiment Setup

The U-Net model is implemented using a deep learning framework (i.e TensorFlow). The model is trained on noisy dataset, with experiments conducted on normalized versions of the data.

4.1.1 Loss Functions

Binary Cross Entropy (BCE): Measures the difference between the predicted probability and actual binary label.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{4.1}$$

Binary Focal Cross Entropy: Focuses on hard-to-classify examples by down-weighting easy ones.

$$FocalLoss = -\frac{1}{N}\sum_{i=1}^{N}(1 - \hat{y}_i)^{\gamma}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{4.2}$$

4.1.2 Activation Functions

ReLU (Rectified Linear Unit): Activates neurons by outputting the input directly if it is positive, otherwise, it outputs zero.

$$ReLU(x) = \max(0, x) \tag{4.3}$$

Tanh: Maps input values to the range $[-1, 1]$ and is symmetric around the origin.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4.4}$$

4.1.3 Epochs

Explanation: One complete pass through the training dataset.

4.1.4 Optimizers

Adam: Adaptive learning rate optimization algorithm that combines the benefits of AdaGrad and RMSProp.

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{4.5}$$

.

SGD (Stochastic Gradient Descent): Simple optimization algorithm that updates parameters by moving in the direction of the negative gradient.

$$\theta_t = \theta_{t-1} - \eta \nabla_\theta J(\theta) \tag{4.6}$$

4.2 Model Uncertainty Reduction

Hyperparameter tuning is performed to reduce model uncertainty. Various configurations are tested, including different activation functions, numbers of hidden layers, and loss functions. The best-performing model is selected based on the evaluation metrics.

Hyperparameter Tuning Using Bayesian Optimization with W&B Sweeps

Algorithm:

1. **Setup W&B Project**:
   - Start by setting up a project in W&B where all the experiment runs will be tracked.
   - Ensure you have the wandb Python package installed and authenticated to your account.
2. **Define the Configuration**:
   - Create a YAML or Python configuration file that specifies the hyperparameters and their ranges, as well as the sweep method.
3. **Initialize the Sweep**:
   - Use the W&B API to initialize the sweep with the defined configuration.
4. **Define the Training Function**:
   - Create a training function that uses the hyperparameters from W&B and logs the results back to W&B.
5. **Run the Sweep**:
   - Launch the sweep agent which will manage the execution of the different runs.
6. **Monitor and Analyze Results**:
   - Use the W&B dashboard to monitor the progress of the sweep in real-time.
   - Analyze the results to find the best hyperparameters based on the evaluation metric.

Chapter 5

Result & Discussion

5.1 Non-Noisy Data Results

The table illustrates the performance of various neural network models on normalized non-noisy data. This section analyzes the effect of different optimizers, activation functions, and loss functions on the overall performance of the models. The results highlight key metrics, including accuracy, AUC score, loss, mean IoU, and precision, which are crucial for evaluating model performance.

ID	Optimiser	Activation	Loss	epoch	accuracy	AUC_score	loss	mean_iou	precision
1	SGD	relu	Binary_focal_cross_entropy	99	0.755	0.573	0.141	0.003	0.659
2	SGD	relu	Binary_cross_enthropy	99	0.751	0.617	0.581	0.000	0.751
3	SGD	tanh	Binary_cross_enthropy	99	0.792	0.815	0.447	0.281	0.671
4	SGD	tanh	Binary_focal_cross_entropy	99	0.752	0.694	0.132	0.131	0.505
5	Adam	tanh	Binary_focal_cross_entropy	99	0.856	0.915	0.081	0.564	0.700
6	Adam	tanh	Binary_cross_enthropy	99	0.867	0.926	0.295	0.571	0.743
7	Adam	relu	Binary_focal_cross_entropy	99	0.907	0.965	0.053	0.692	0.794
8	Adam	relu	Binary_cross_enthropy	99	0.912	0.968	0.195	0.693	0.844

Figure 5.1: Performance Metrics on Normalized Non-Noisy Data

Table Explanation

ID: This is a unique identifier for each experiment conducted. It helps in distinguishing between the various configurations tested.

Optimizer: This column specifies the optimization algorithm used to minimize the loss function. The optimizers used in this table are:

- SGD (Stochastic Gradient Descent)
- Adam (Adaptive Moment Estimation)

Activation: This column lists the activation function applied to the neurons in the network. Activation functions help in introducing non-linearity into the model, enabling it to learn complex patterns. The activation functions used are:

- ReLU (Rectified Linear Unit)
- Tanh (Hyperbolic Tangent)

Loss: This column indicates the loss function used to compute the error during training. Two types of loss functions are compared:

- Binary Focal Cross-Entropy
- Binary Cross-Entropy

Epoch: This represents the number of complete passes through the training dataset. Here, all experiments were conducted for 99 epochs.

Metrics: **Accuracy:** It measures the percentage of correct predictions made by the model out of all predictions.

AUC Score (Area Under the Curve): This metric evaluates the ability of the model to distinguish between classes. Higher AUC scores indicate better model performance.

Loss: This is the value of the loss function after training. Lower loss values typically indicate better performance.

Mean IoU (Intersection over Union): It measures the overlap between the predicted and true segments. Higher mean IoU values indicate better performance.

Precision: This measures the proportion of true positive predictions among all positive predictions. Higher precision indicates fewer false positives.

Analysis of Results

1. **Optimizer Comparison:**

   - The Adam optimizer generally outperforms SGD in terms of accuracy, AUC score, and mean IoU, as seen from the higher values in rows 5-8 compared to rows 1-4.
   - Adam also shows lower loss values, indicating more effective minimization of the loss function.

2. **Activation Function Comparison:**

   - Comparing ReLU and Tanh within the same optimizer and loss function combinations, Tanh (rows 3-4) often results in slightly higher accuracy and AUC scores than ReLU (rows 1-2) for SGD.
   - For Adam, ReLU (rows 7-8) demonstrates higher performance metrics than Tanh (rows 5-6), suggesting a better fit for the normalized non-noisy data.

3. **Loss Function Impact:**

   - Models using the Binary Cross-Entropy loss function (rows 2, 4, 6, 8) consistently show higher accuracy, AUC scores, mean IoU, and precision compared to those using the Binary Focal Cross-Entropy loss function (rows 1, 3, 5, 7).

4. **Overall Best Performance:**

   - The best performing model is in row 8 (Adam optimizer, ReLU activation, Binary Cross-Entropy loss), achieving the highest accuracy (0.912), AUC score (0.968), mean IoU (0.693), and precision (0.844) with a relatively low loss value (0.195).

5.2 Confusion Matrix Results

The table illustrates the confusion matrix results for various neural network models across different datasets: Training Set, Validation Set, and Testing Set. This section analyzes the True Positives (TP), False Positives (FP), and False Negatives (FN) for each dataset to understand the model performance.

ID	Training Set	Validation Set	Testing Set
1	TP : 544 FP : 336 FN : 1798759	TP : 16 FP : 63 FN : 311985	TP : 598 FP : 361 FN : 1987697
2	TP : 0 FP : 2 FN : 1823349	TP : 0 FP : 0 FN : 139153	TP : 0 FP : 2 FN : 1988295
3	TP : 721546 FP : 378238 FN : 1101803	TP : 16879 FP : 94059 FN : 122274	TP : 801094 FP : 409628 FN : 1187201
4	TP : 141412 FP : 94209 FN : 1681937	TP : 3944 FP : 23419 FN : 135209	TP : 160819 FP : 102065 FN : 1827376
5	TP : 1418952 FP : 590942 FN : 404397	TP : 27732 FP : 135186 FN : 111421	TP : 1539018 FP : 633794 FN : 449277
6	TP : 334362 FP : 76279 FN : 1488987	TP : 7934 FP : 43547 FN : 131219	TP : 375931 FP : 86191 FN : 1612364
7	TP : 1334292 FP : 213259 FN : 489057	TP : 23481 FP : 116601 FN : 115672	TP : 1454827 FP : 232806 FN : 533468
8	TP : 1660432 FP : 498512 FN : 162917	TP : 30231 FP : 149486 FN : 108922	TP : 1798580 FP : 540081 FN : 189715

Figure 5.2: Confusion Matrix Results for Training, Validation, and Testing Sets

Table Explanation

ID: This is a unique identifier for each experiment conducted. It helps in distinguishing between the various configurations tested.

Training Set: This column shows the TP, FP, and FN values for the training dataset.

Validation Set: This column presents the TP, FP, and FN values for the validation dataset.

Testing Set: This column provides the TP, FP, and FN values for the testing dataset.

Analysis of Results

1. **Training Set:**
   - Most models show a high number of False Negatives (FN), indicating that many true instances were missed.
   - The number of False Positives (FP) varies significantly, with some models having very low FP values (e.g., ID 2) and others having higher values.
2. **Validation Set:**
   - The validation set shows similar trends to the training set, with a high number of FNs.
   - Models generally have fewer TP and FP values in the validation set compared to the training set, reflecting the smaller size of the validation dataset.
3. **Testing Set:**
   - The testing set results are crucial for evaluating the generalization performance of the models.
   - Similar to the training and validation sets, the testing set also shows a high number of FNs.
   - The FP values are relatively lower in the testing set compared to the training set, indicating improved precision in some models (e.g., ID 3, 4).
4. **Overall Best Performance:**
   - The best performing model based on TP values in the testing set appears to be ID 8, with a TP value of 540081.
   - ID 8 also has the lowest FN value in the testing set, indicating better recall.

5.3 Graphs for Non-Noisy Data:

Validation Metrics (Accuracy and Loss)

Implications:

- The high and stable validation accuracy suggests good generalization to the validation set.

5.3.1 General Description of Training and Validation Metrics

The graphs presented below illustrate the training and validation metrics, specifically accuracy and loss, over the epochs of model training. These metrics are crucial for evaluating the performance and generalization ability of the model.

- **Training Accuracy vs. Epoch**: This graph shows the accuracy of the model on the training dataset over the number of epochs. A consistently increasing training accuracy indicates that the model is learning from the training data effectively.
- **Validation Accuracy vs. Epoch**: This graph depicts the accuracy of the model on the validation dataset. A high and stable validation accuracy is a sign that the model is generalizing well to new, unseen data.
- **Training Loss vs. Epoch**: This graph displays the loss on the training dataset. A decreasing training loss suggests that the model is fitting the training data well.
- **Validation Loss vs. Epoch**: This graph shows the loss on the validation dataset. A decreasing validation loss is desirable as it indicates that the model is not just memorizing the training data but also performing well on new data.

These are graphs for respective IDs runs.

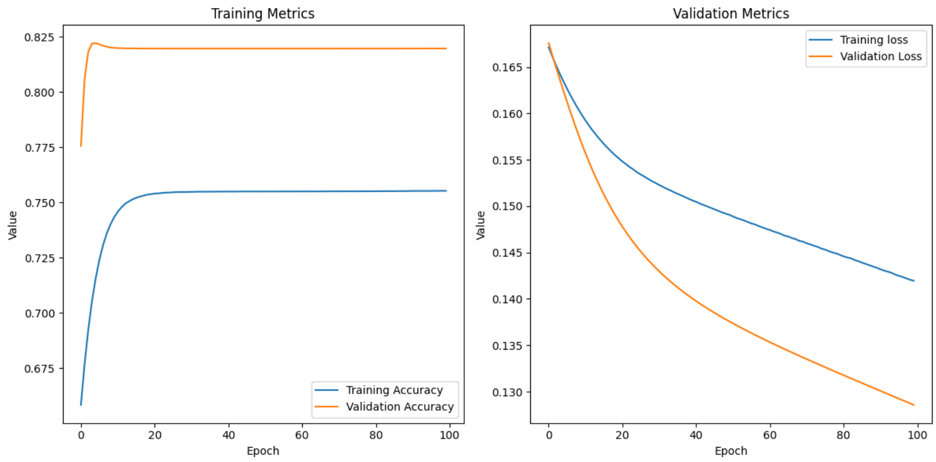

Figure 5.3: Accuracy vs loss graph-1(SGD ReLU Binary Focal Cross Entropy)

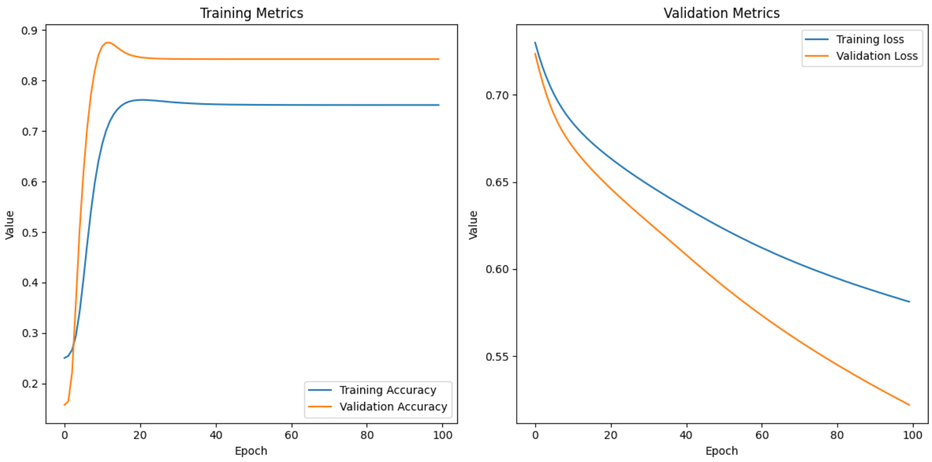

Figure 5.4: Accuracy vs loss graph-2(SGD ReLU Binary Cross Entropy)

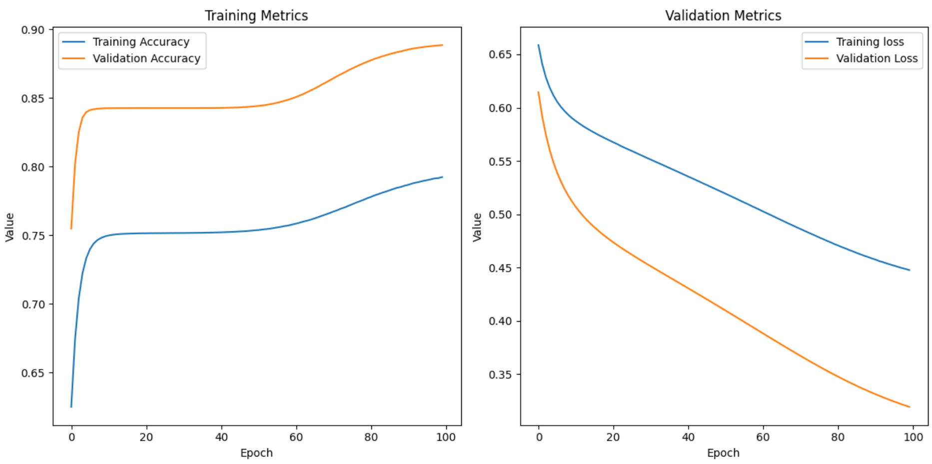

Figure 5.5: Accuracy vs loss graph-3(SGD Tanh Binary Cross Entropy)

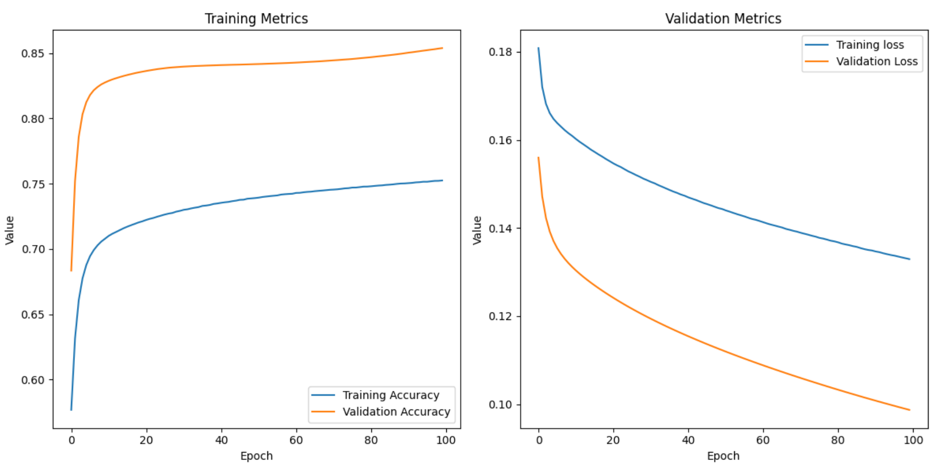

Figure 5.6: Accuracy vs loss graph-4(SGD Tanh Binary Focal Cross Entropy)

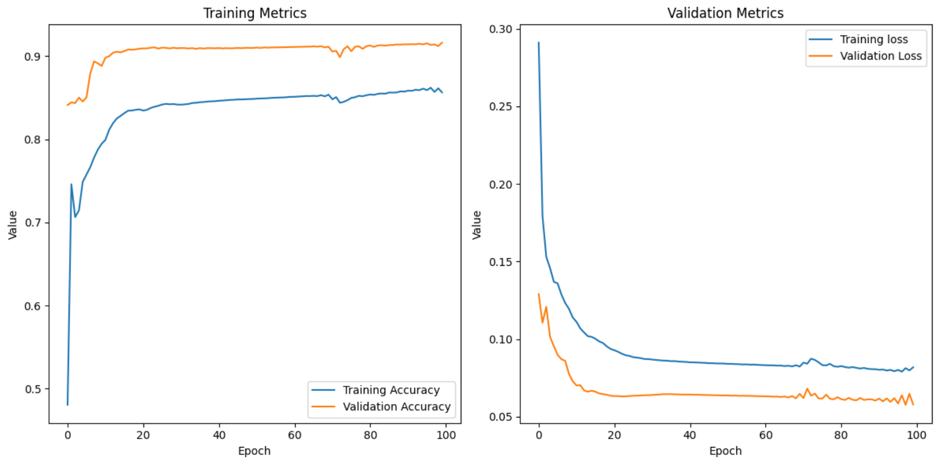

Figure 5.7: Accuracy vs loss graph-5(Adam Tanh Binary Focal Cross Entropy)

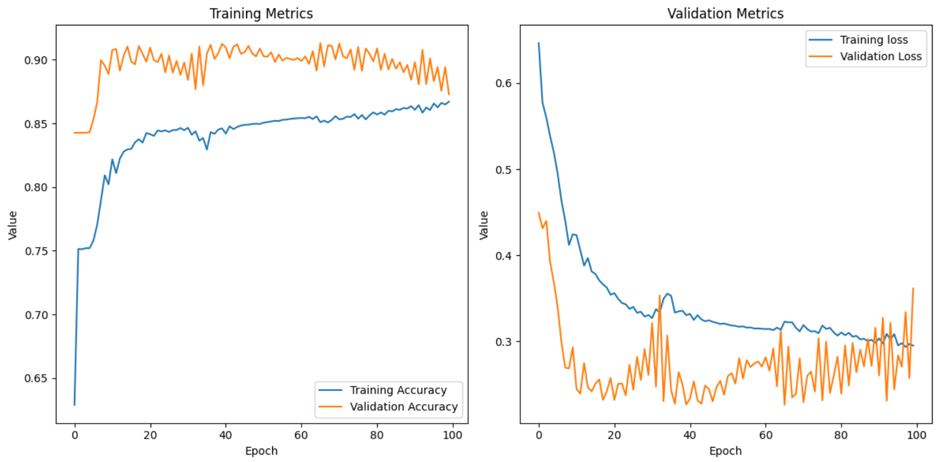

Figure 5.8: Accuracy vs loss graph-6(Adam Tanh Binary Cross Entropy)

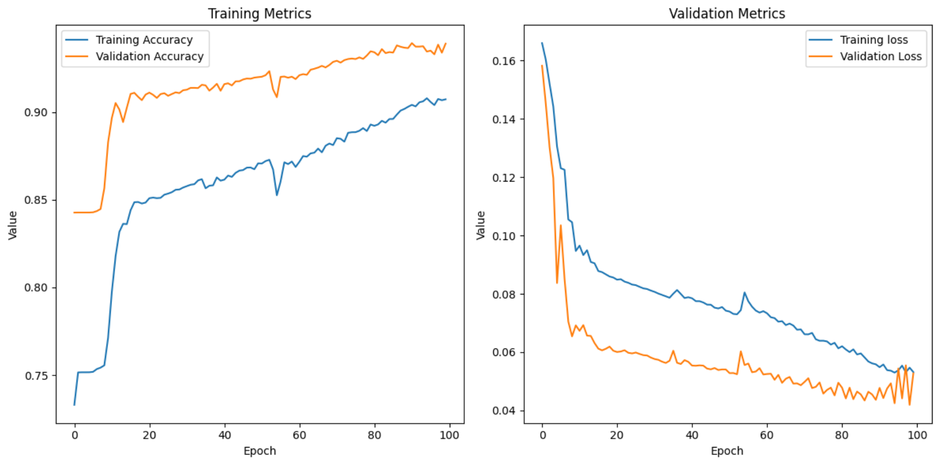

Figure 5.9: Accuracy vs loss graph-7(Adam ReLU Binary Focal Cross Entropy)

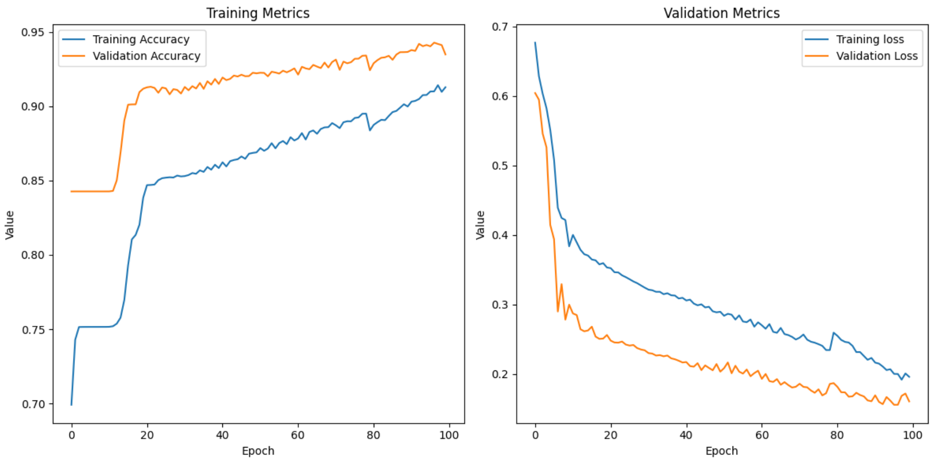

Figure 5.10: Accuracy vs loss graph-8(Adam ReLU Binary Cross Entropy)

5.3.2 Predictions

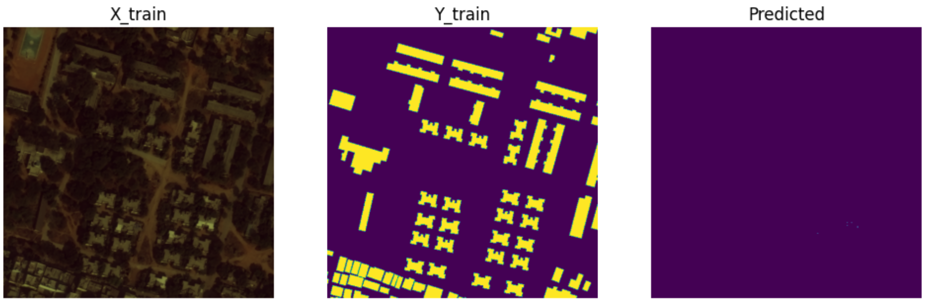

Figure 5.11: Prediction-1(SGD ReLU Binary Focal Cross Entropy)

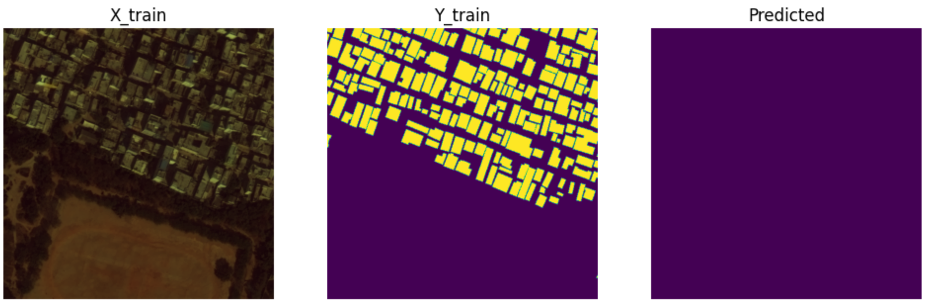

Figure 5.12: Prediction-2(SGD ReLU Binary Cross Entropy)

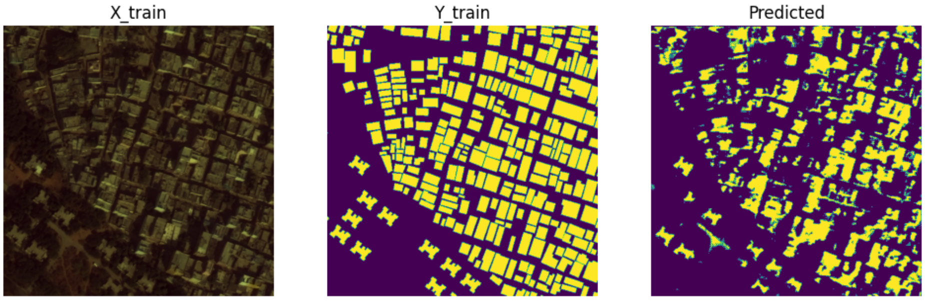

Figure 5.13: Prediction-3(SGD Tanh Binary Cross Entropy)

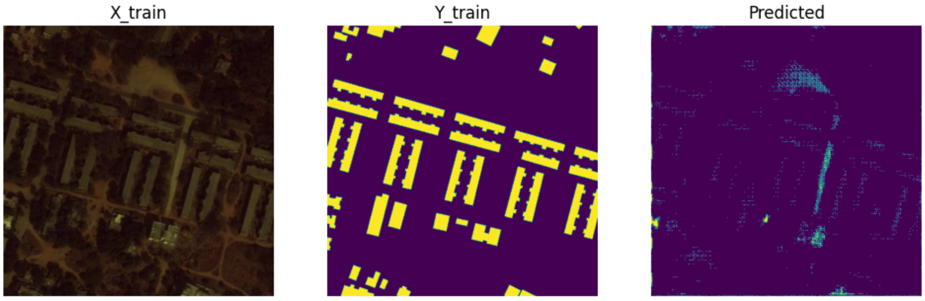

Figure 5.14: Prediction-4(SGD Tanh Binary Focal Cross Entropy)

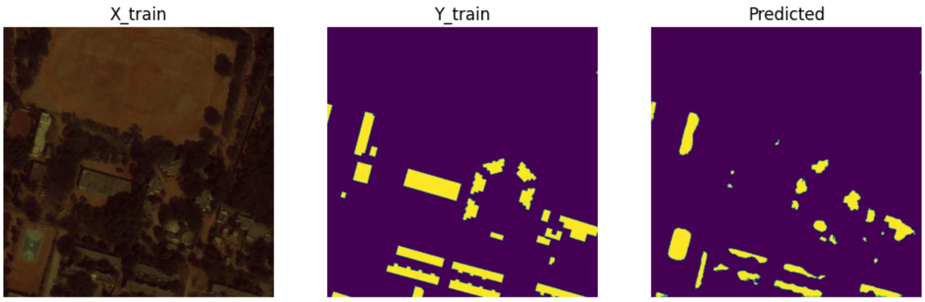

Figure 5.15: Prediction-5(Adam Tanh Binary Focal Cross Entropy)

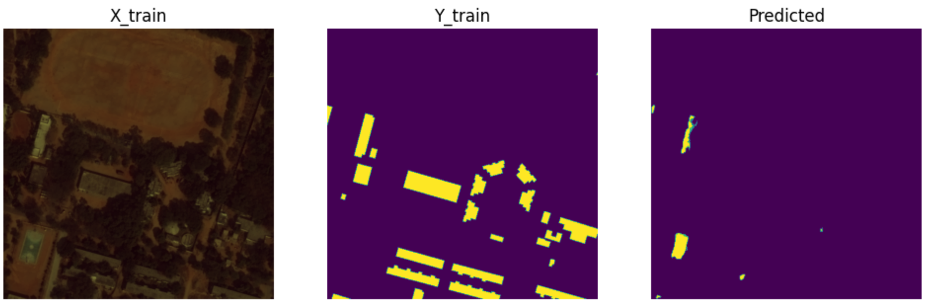

Figure 5.16: Prediction-6(Adam Tanh Binary Cross Entropy)

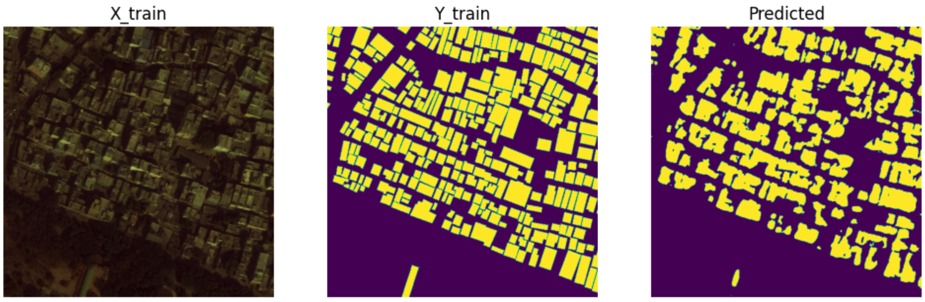

Figure 5.17: Prediction-7(Adam ReLU Binary Focal Cross Entropy)

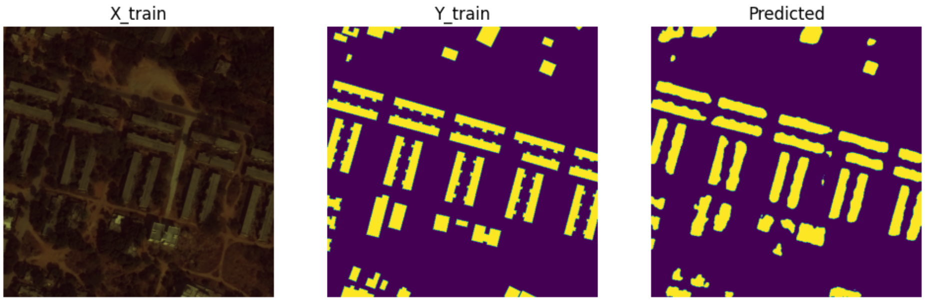

Figure 5.18: Prediction-8(Adam ReLU Binary Cross Entropy)

5.3.3 Confidence Masks

Introduction

Confidence masks are an essential tool in image segmentation and object detection tasks within the field of computer vision. They provide a way to visualize the certainty of the model's predictions and help in assessing the quality of the segmentation or detection results.

Concept of Confidence Masks

In the context of image segmentation, a confidence mask represents the model's confidence in predicting the presence of a particular object or class in each pixel of the image. Each pixel in the confidence mask is associated with a confidence score, typically ranging from 0 to 1, where:

- **0**: Indicates no confidence in the presence of the object.
- **1**: Indicates full confidence in the presence of the object.

These masks are particularly useful for understanding and visualizing the areas where the model is certain about its predictions versus areas where it is uncertain.

Visualization and Interpretation

The figure below illustrates examples of images, their corresponding masks, and the combined visualization with confidence scores:

- **Image**: The original input image on which segmentation is performed.

- **Mask**: The binary mask indicating the predicted regions of interest, such as buildings or objects.
- **Image with Mask and Confidence**: The combined visualization showing the original image overlaid with the predicted mask and confidence scores.

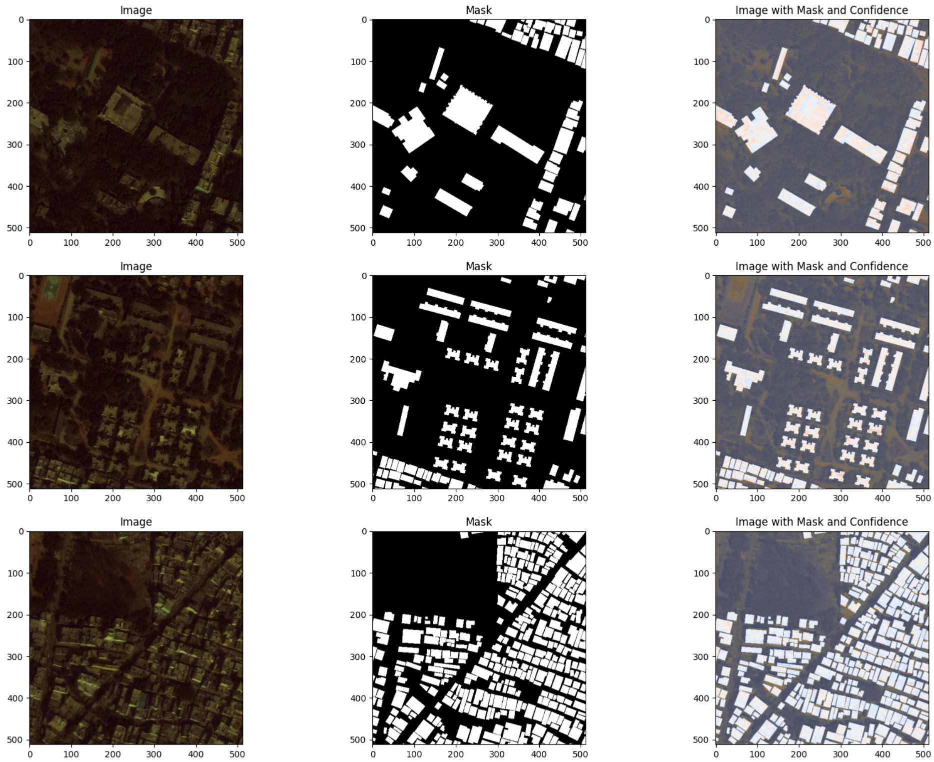

Figure 5.19: Confidence Score-1(SGD ReLU Binary Focal Cross Entropy)

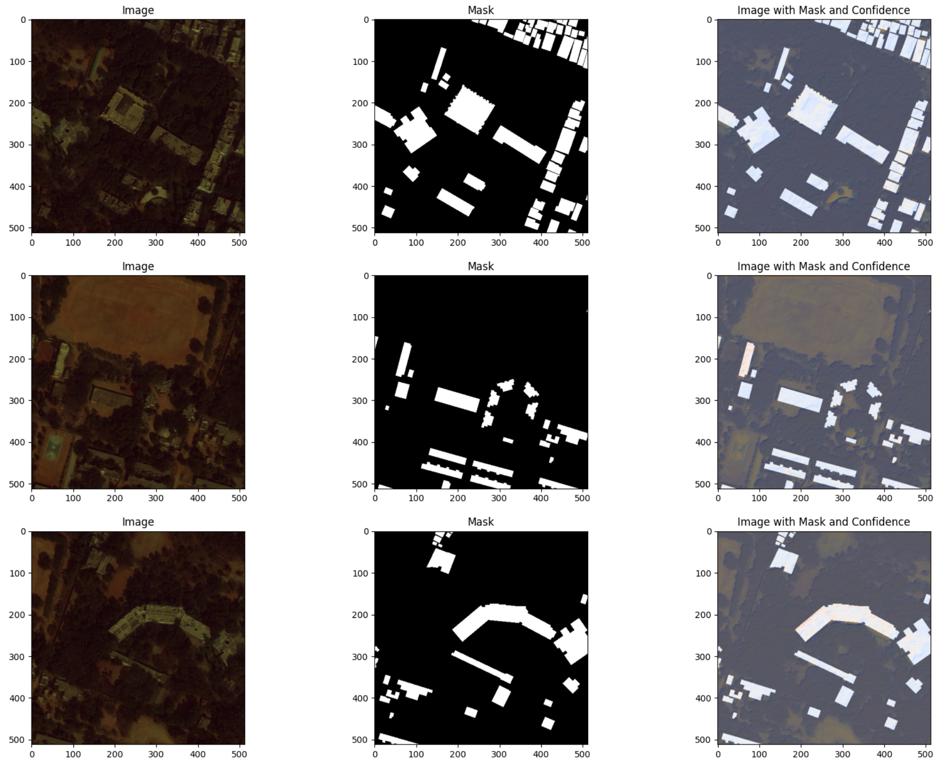

Figure 5.20: Confidence Score-2(SGD ReLU Binary Cross Entropy)

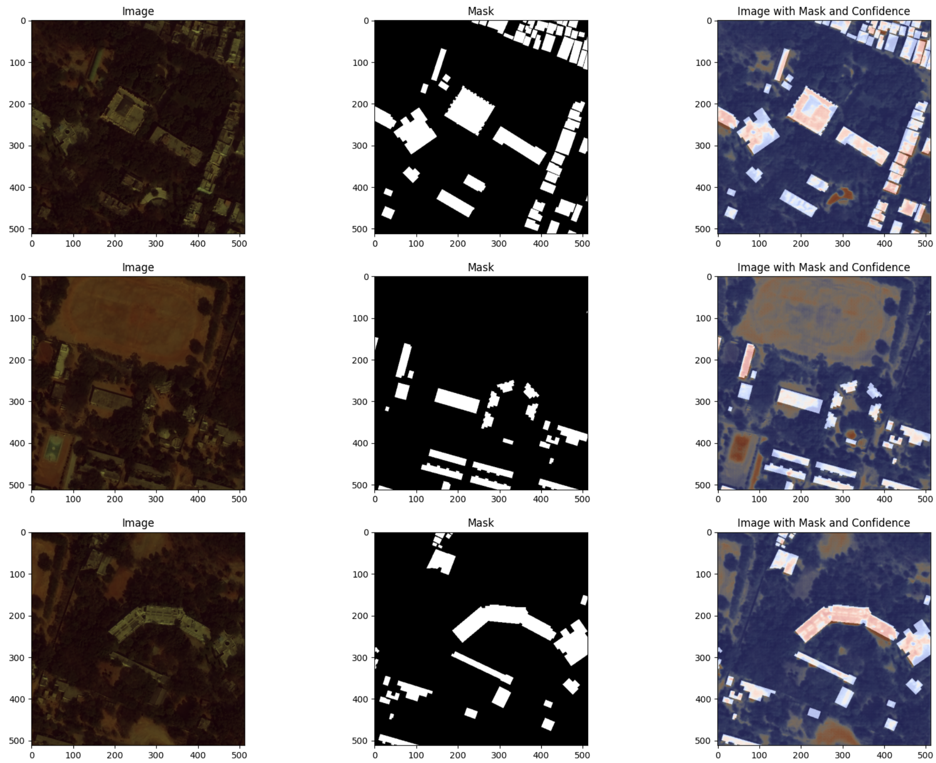

Figure 5.21: Confidence Score-3(SGD Tanh Binary Cross Entropy)

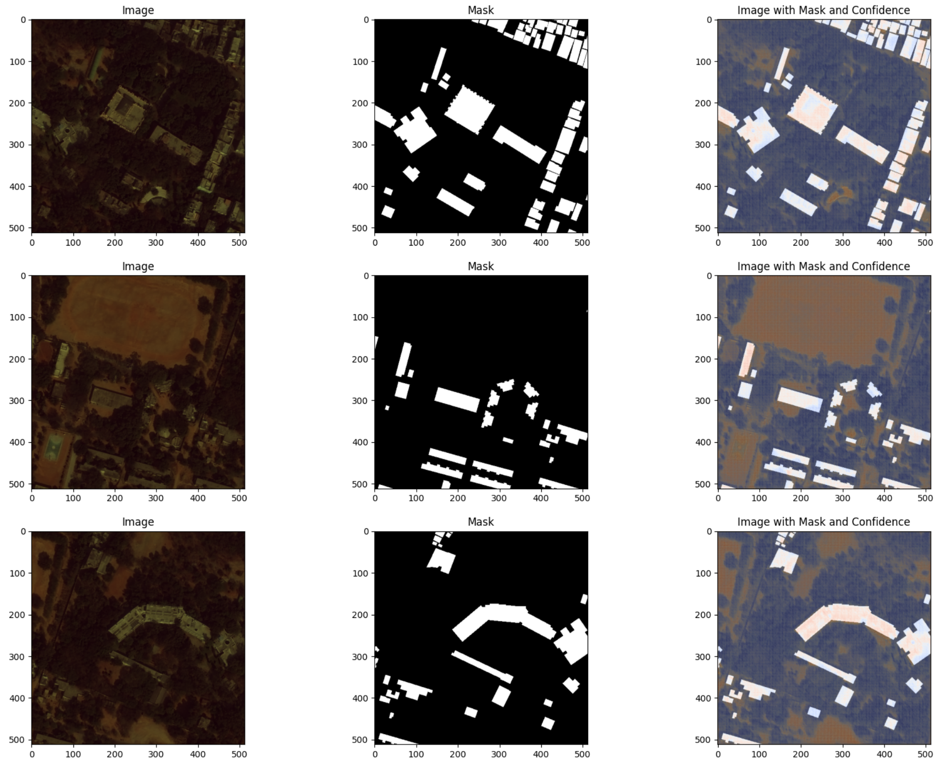

Figure 5.22: Confidence Score-4(SGD Tanh Binary Focal Cross Entropy)

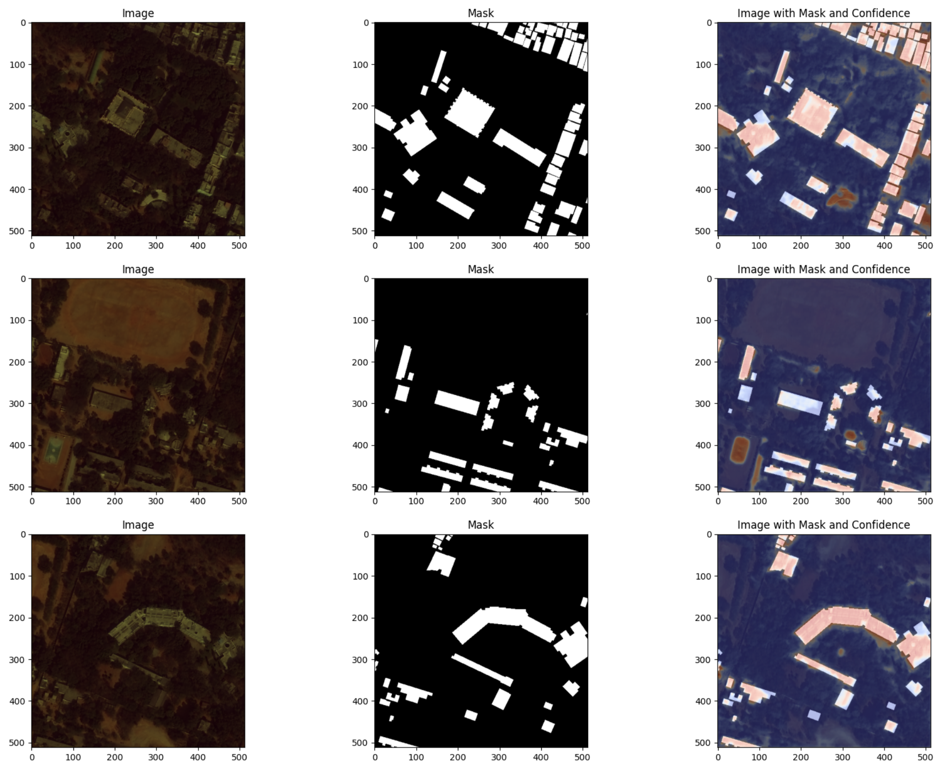

Figure 5.23: Confidence Score-5(Adam Tanh Binary Focal Cross Entropy)

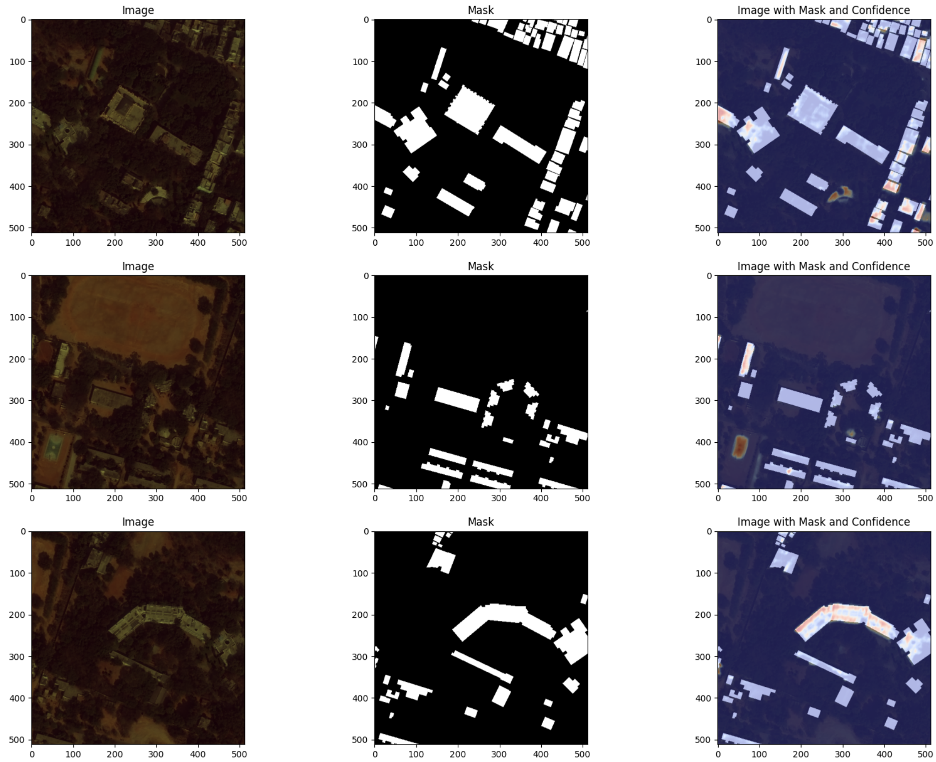

Figure 5.24: Confidence Score-6(Adam Tanh Binary Cross Entropy)

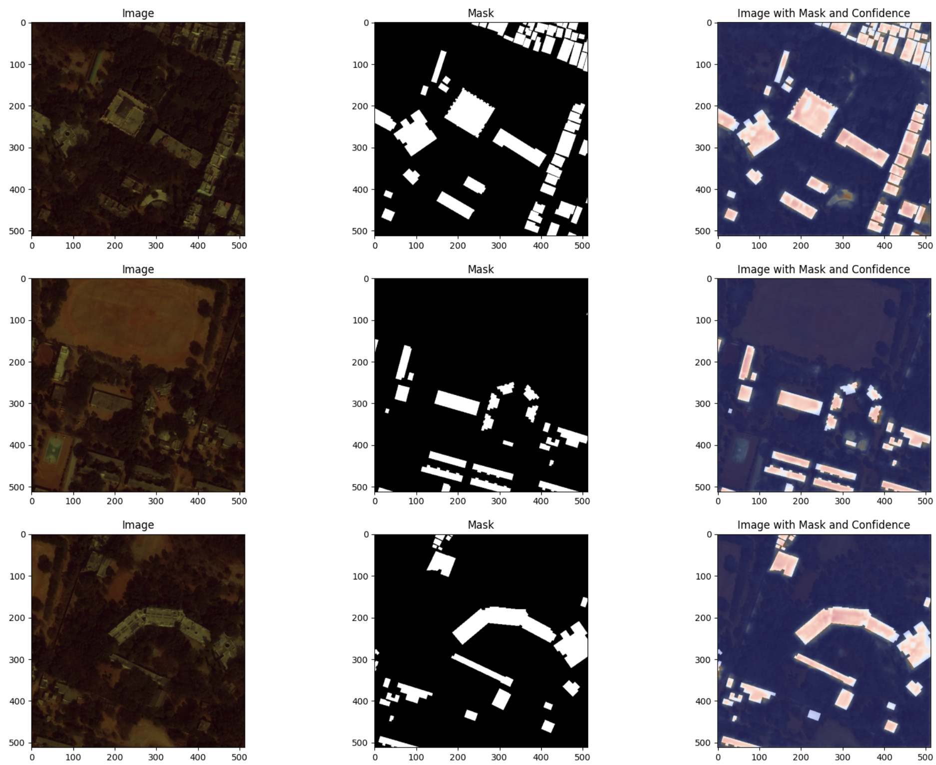

Figure 5.25: Confidence Score-7(Adam ReLU Binary Focal Cross Entropy)

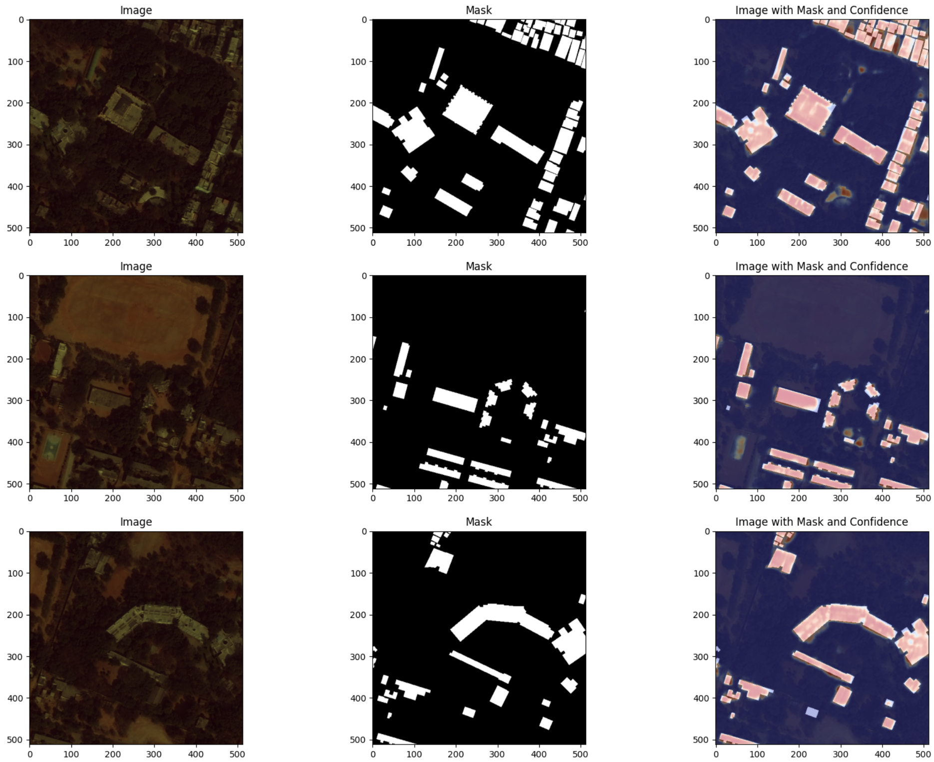

Figure 5.26: Confidence Score-8(Adam ReLU Binary Cross Entropy)

Importance of Confidence Masks

Confidence masks play a crucial role in evaluating the performance of segmentation models. They offer several benefits:

- **Model Validation**: By visualizing confidence masks, one can validate how confident the model is in its predictions, which is crucial for applications requiring high reliability.
- **Error Analysis**: Confidence masks help in identifying areas where the model is uncertain, providing insights into potential areas for improvement or further training.
- **Decision Making**: In practical applications, such as autonomous driving or medical imaging, confidence scores can guide decision-making processes by indicating the level of certainty in the model's predictions.

Chapter 6

Conclusion

This report presents a comprehensive analysis of neural network models trained and evaluated on normalized non-noisy data, with an emphasis on model performance metrics and the use of confidence masks in image segmentation tasks.

Performance Analysis

Through various experiments, we identified key factors contributing to optimal model performance:

- **Optimizer Effectiveness:** The Adam optimizer consistently outperforms SGD, yielding higher accuracy, AUC score, mean IoU, and precision.
- **Activation Function Comparison:** ReLU activation, in conjunction with the Adam optimizer, demonstrates superior performance metrics compared to Tanh.
- **Loss Function Impact:** Models employing the Binary Cross-Entropy loss function outperform those using Binary Focal Cross-Entropy, indicating better optimization and generalization.

Model ID 8, utilizing the Adam optimizer, ReLU activation, and Binary Cross-Entropy loss function, achieves the highest performance across various metrics, including accuracy, AUC score, mean IoU, and precision, alongside a relatively low loss value. This configuration emerges as the optimal choice for the given dataset.

Confusion Matrix Insights

The confusion matrix results highlight:

- **High False Negatives:** There is a notable presence of False Negatives across most models, indicating a need for improved detection capabilities.
- **Variability in False Positives:** Some models show very low False Positive rates, reflecting high precision, but variability exists.
- **Best Performance:** Model ID 8 exhibits the best performance in the testing set, with the highest True Positives and lowest False Negatives, indicating strong generalization and recall.

Training and Validation Metrics

The training and validation metrics over epochs reveal important trends:

- **Accuracy Trends:** A steady increase in training accuracy and consistently high validation accuracy suggests effective learning and good generalization.
- **Loss Trends:** Rapid decreases in training and validation loss in early epochs point to effective optimization, although a significant gap between them may indicate potential overfitting.
- **Epoch Analysis:** Early epochs show rapid improvement, with later epochs demonstrating convergence toward optimal performance.

Confidence Masks

Confidence masks provide a detailed visual assessment of the model's prediction certainty in image segmentation tasks. These masks are crucial for:

- **Model Validation and Error Analysis:** By visualizing confidence scores for each pixel, confidence masks help in identifying areas where the model performs well and areas needing further refinement.
- **Decision-Making:** The combined visualizations of images, masks, and confidence scores offer valuable insights for improving model training and decision-making processes.

Overall Implications

The combined insights from the performance metrics, confusion matrix analysis, and confidence mask visualizations provide a holistic view of model evaluation. The Adam optimizer with ReLU activation and Binary Cross-Entropy loss function emerges as the most effective configuration for this dataset, achieving the best overall performance. However, the high number of False Negatives indicates a need for further tuning to enhance model recall. The use of confidence masks complements traditional metrics, offering a comprehensive evaluation approach that ensures reliable and interpretable predictions in practical applications.

This dual approach of combining quantitative metrics with visual assessments ensures robust model evaluation, essential for developing high-quality, reliable machine learning models suitable for real-world applications such as autonomous driving, medical imaging, and other critical domains.

References

[1] Waleed Alsabhan. *Computational Intelligence and Neuroscience. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net,* Volume 2022 — Article ID 4831223 .

[2] S.Aghayari, A.Hadavand, S.Mohamadnezhad Niazi, M.Omidalizarandi. *Building detection from aerial imagery using inception resnet unet and unet architectures* . Volume X-4/W1-2022, 19–22 February 2023.

[3] Isah Charles Saidu and Lehel Csató,
Active Learning with Bayesian UNet for Efficient Semantic Image Segmentation) Volume 2021 Feb; 7(2): 37., 2021

[4] Clément Dechesne; Pierre Lassalle; Sébastien Lefèvre,
Bayesian Deep Learning with Monte Carlo Dropout for Qualification of Semantic Segmentation) Published in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS

[5] Totem,
Hyperparameter Tuning in Deep Learning-Based Image Classification to Improve Accuracy using Adam Optimization) Vol. 19 Issue (9): 579-586.2023

[6] Weights Biases Documentation,
https://docs.wandb.ai/