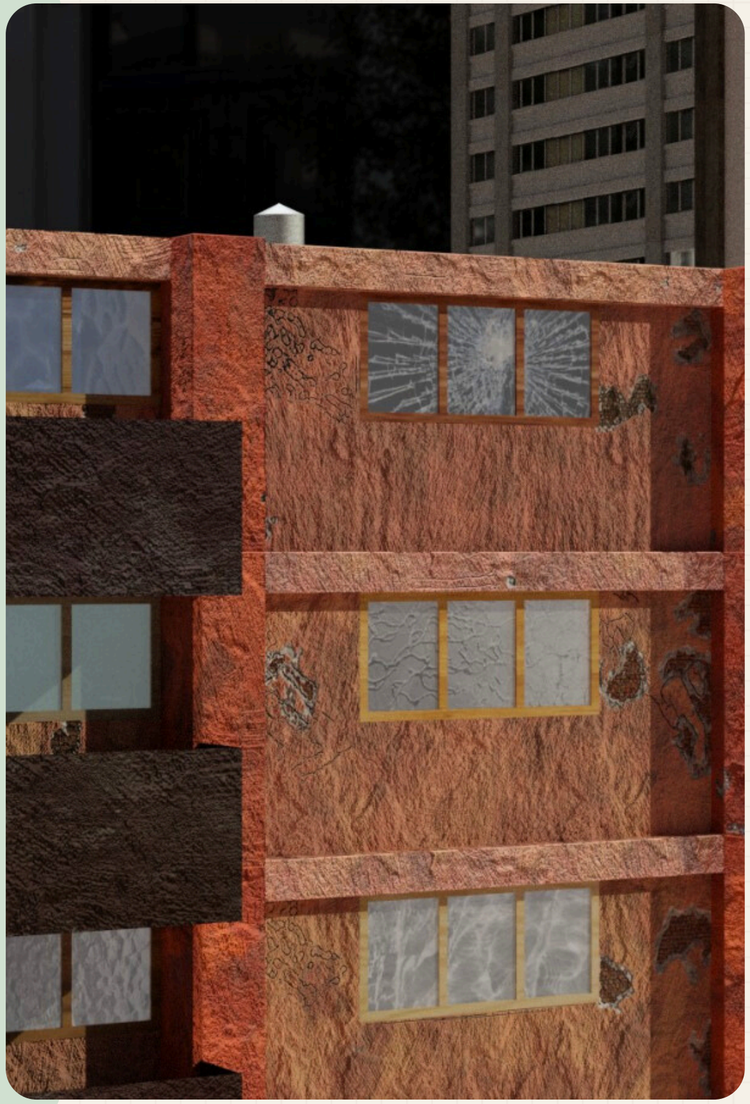
LARGE-SCALE SYNTHETIC DATASET FOR VISION-BASED STRUCTURAL CONDITION ASSESSMENT OF BUILDING FACADES

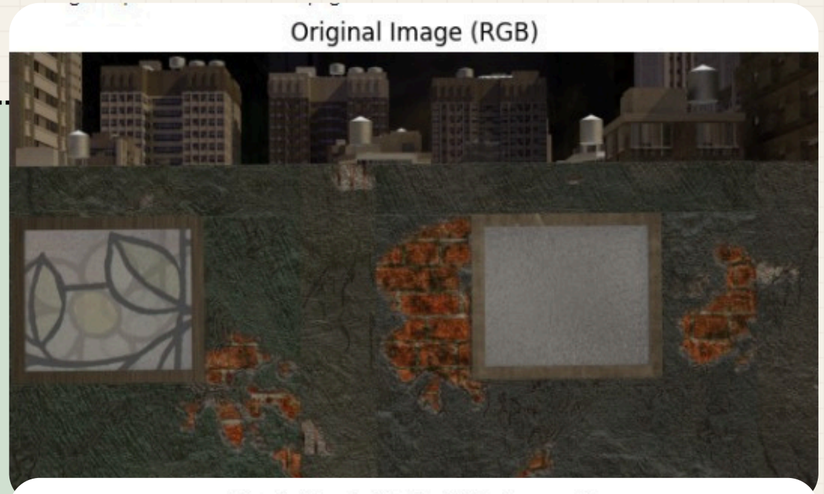
Using Multiclass Semantic Segmentation with U-Net and LSTM

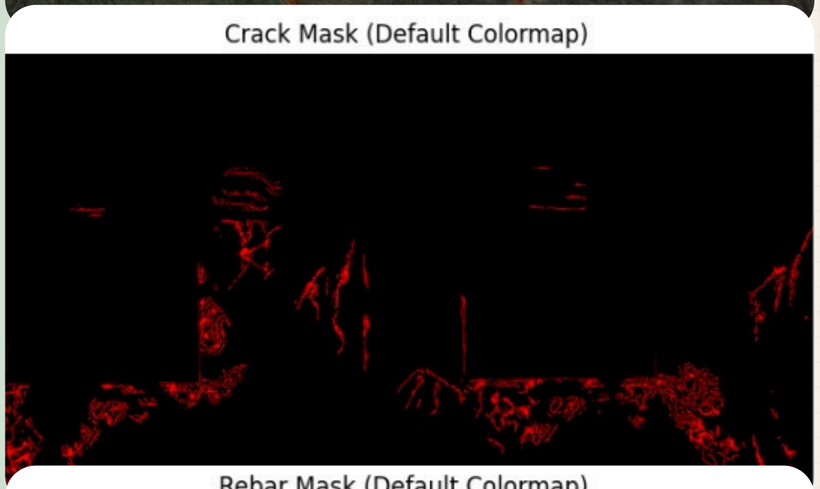


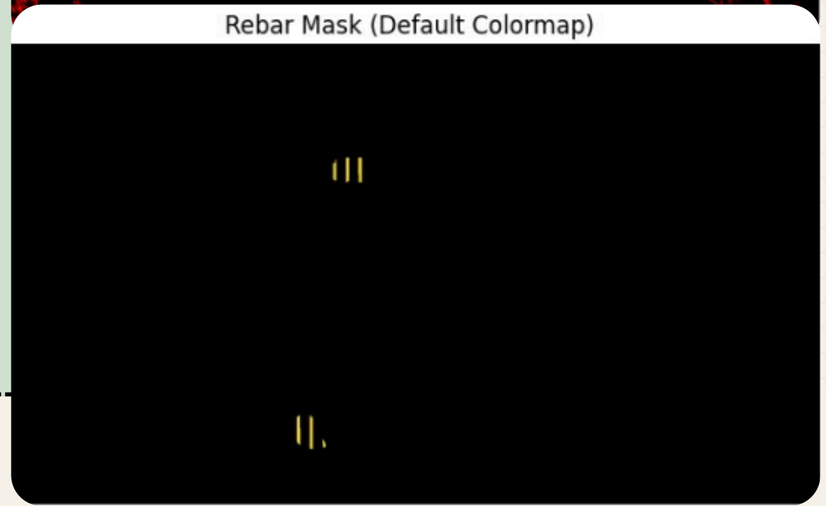
Presented by:

Aditya Prabhakar, Kartik Jaiswal, G Anamika

CONTENT







INTRODUCTION

Develop a deep learning framework capable of identifying and classifying different structural conditions from synthetic visual data of an urban structure

Challenges	Approach
Limited availability of labeled real-world data	Use synthetic dataset mimicking realistic structures
Cracks are often tiny and can be missed by downsampling layers	Apply U-Net for pixel-level segmentation
Small, subtle, and irregular defects are hard to detect	ConvLSTM adds recurrence in the encoder, letting the model iteratively refine features

DATA DESCRIPTION

Data Set Type

Synthetic image dataset of urban structures

Data Source

QuakeCity – Synthetic dataset of earthquake damaged buildings

11 Classes

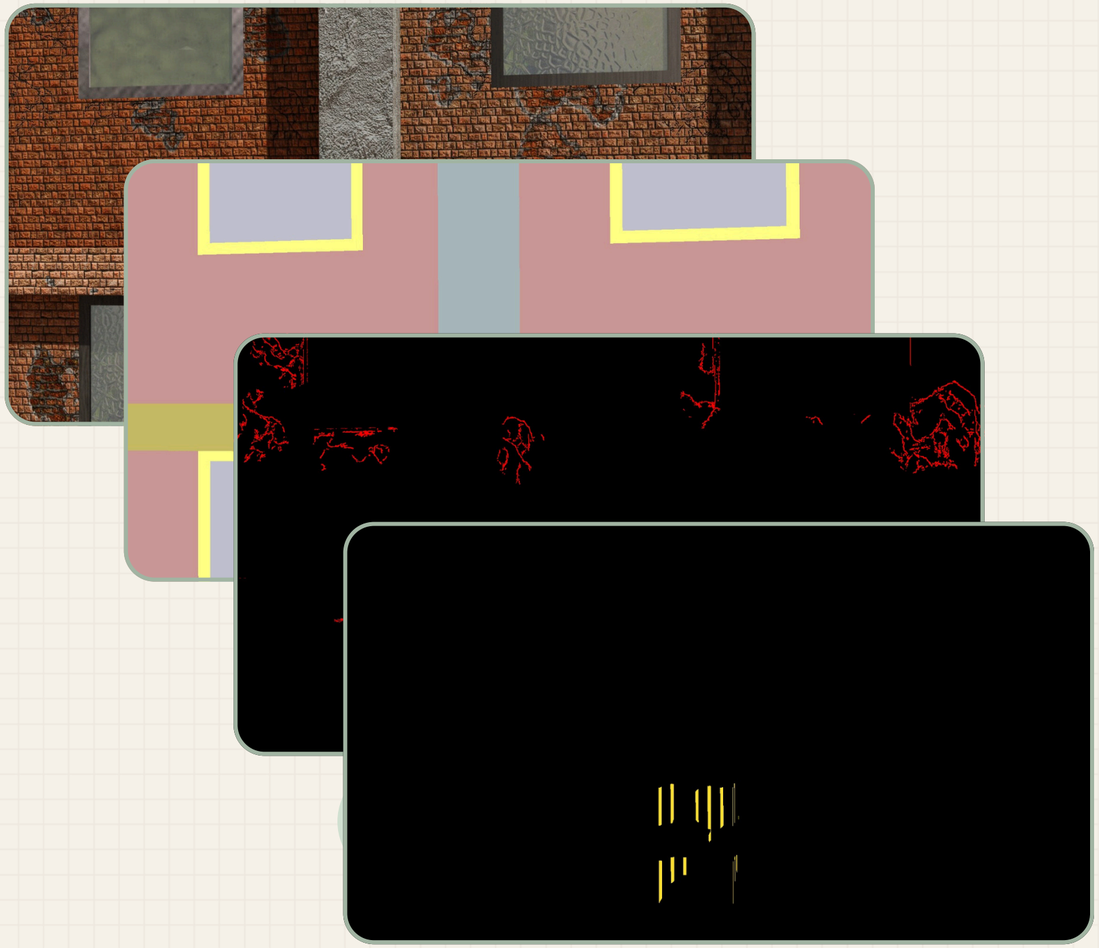
0: Background (BG); 1: Crack; 2: Rebar; 3: Wall;
4: Beam; 5: Column; 6: Window Frame;
7: Window Pane; 8: Balcony; 9: Slab; 10: Ignore

Mask

Combining separate masks into One single mask

Data Volume

4809 Images



Rebar Mask

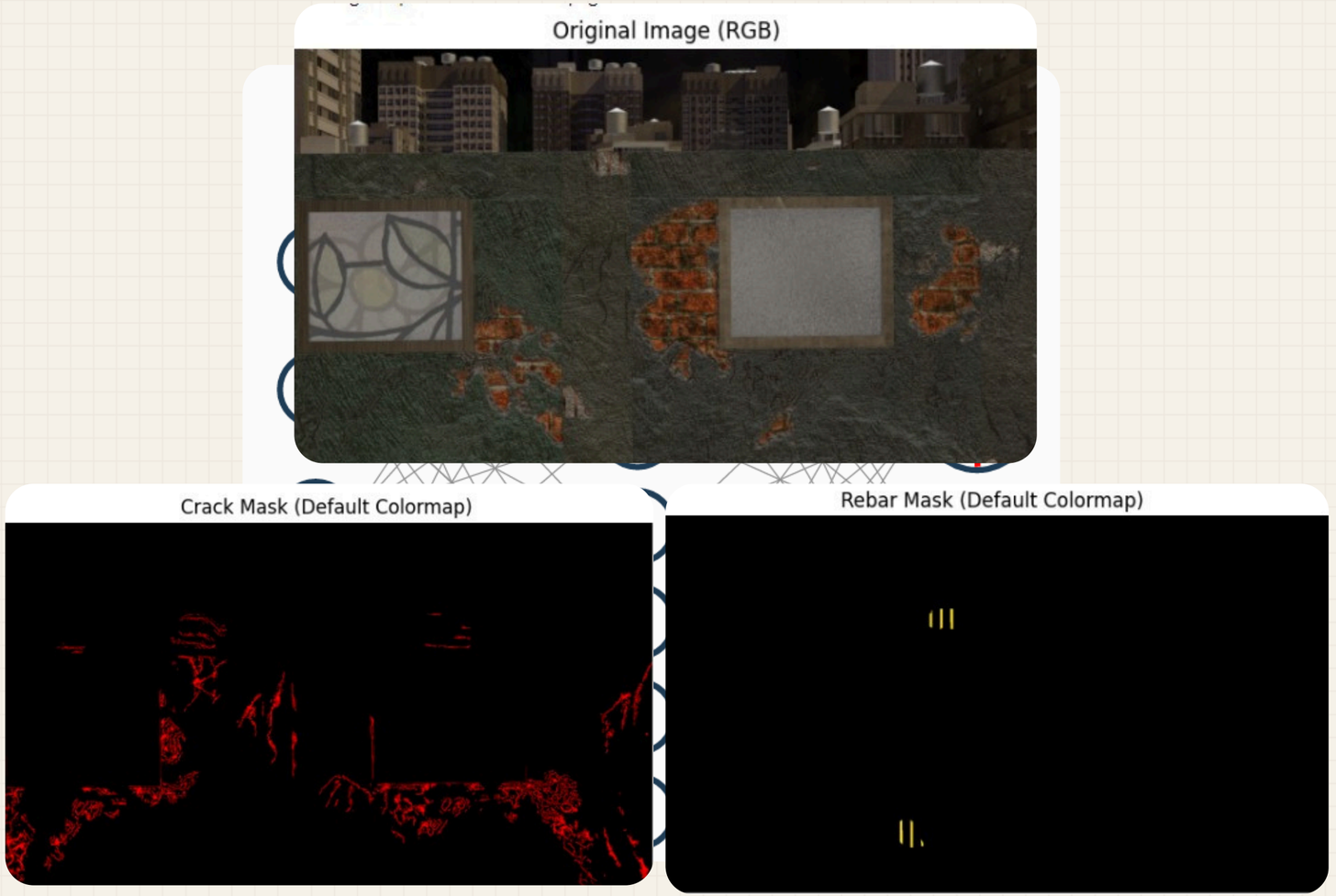
Each pixel in the mask has a label or value that represents the class of that pixel

MULTICLASS SEMANTIC SEGMENTATION

Multiclass distinguishes multiple classes within the same image

Semantic Segmentation builds back the image identify what and where different objects or regions are in the image

It assigns a class label to every pixel in the image



Multilclass Semantic Segmentation

The same model has identified and displayed the cracks and exposed rebars

METHODOLOGY

Data Preparation

Obtaining synthetic dataset & preprocessing

Model Architecture

U-Net for spatial segmentation + LSTM for temporal learning

Training Strategy

Data split, Dice Loss Function, Adam optimization

Evaluation Metrics

Quantitative and visual analysis

Adam = Adaptive Moment Estimation

DATA PREPARATION

Input Data

Synthetic RGB image sequences representing exterior facades of buildings

Image Resolution

288 x 512
(Reduced from 576 x 1024)

Preprocessing

- Image normalization (0–1 scaling)
- Resizing to uniform dimensions
- **2D** image data (**H, W, C**) into a **5D** tensor(Batch, Time, H, W, C) that a ConvLSTM2D layer requires

Data Split

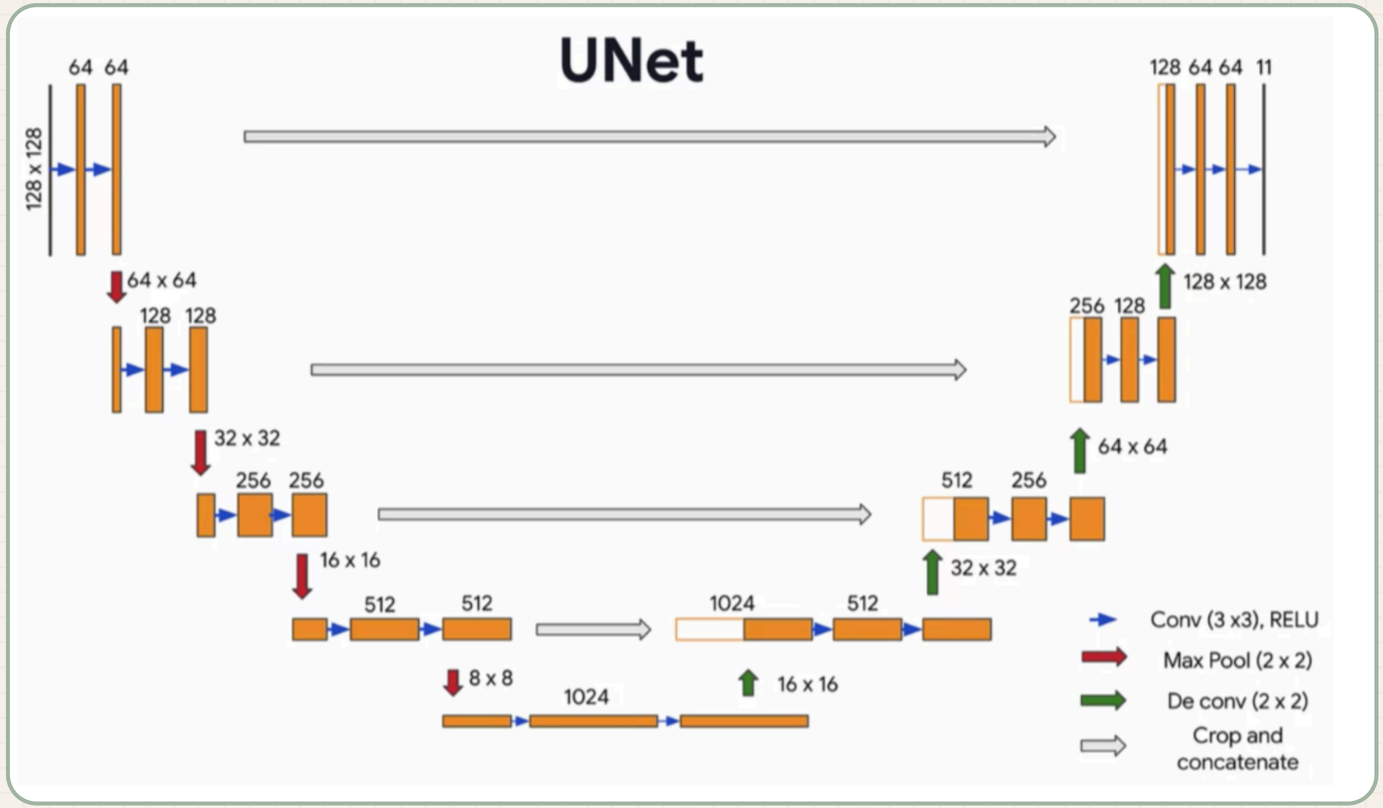
- Training: 3424
- Validation: 381
- Testing: 1004

MODEL ARCHITECTURE

U-Net encoder → Skip Connections → U-Net decoder → Segmentation mask

U-Net

- Encoder–decoder structure ideal for semantic segmentation
- Preserves spatial context through skip connections



U-Net

MODEL ARCHITECTURE

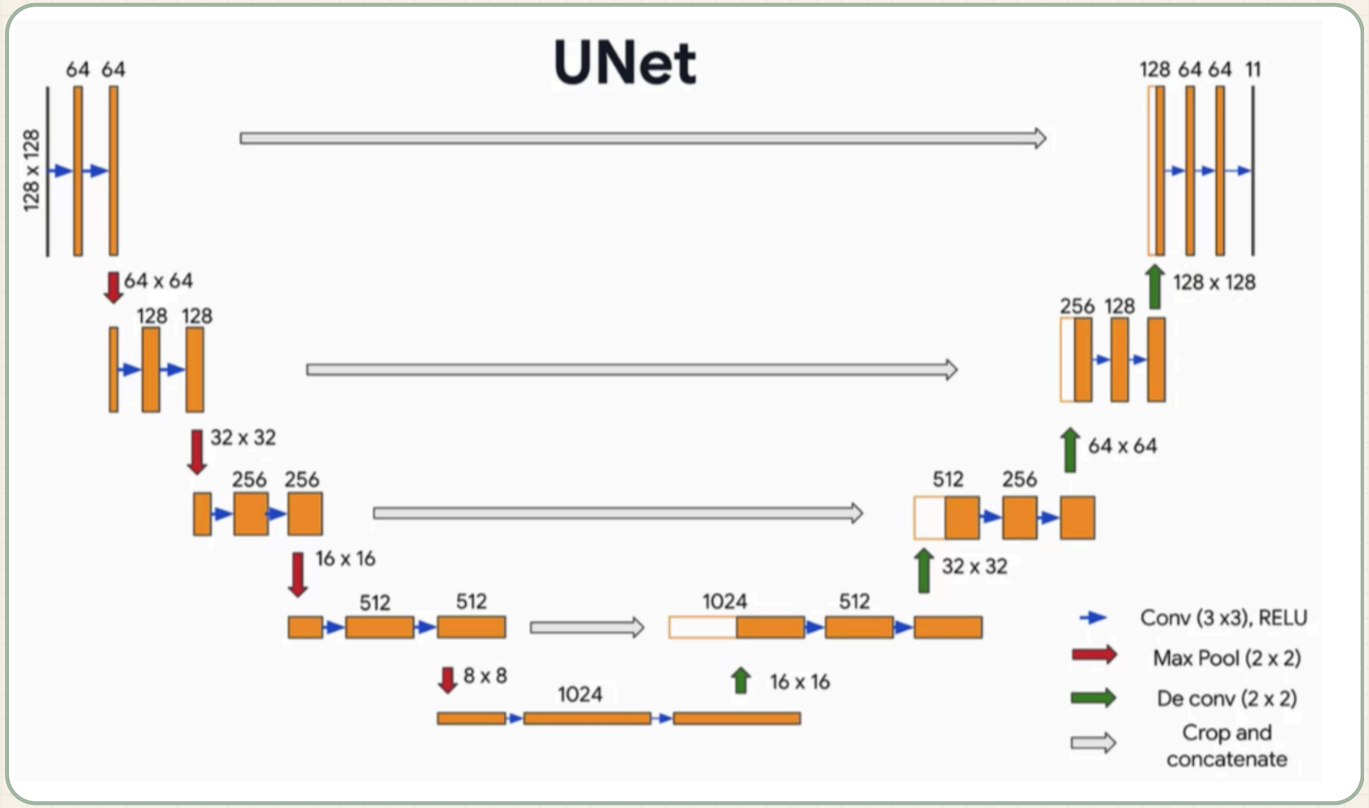
U-Net encoder → convLSTM → U-Net decoder → Segmentation mask

U-Net

- Encoder–decoder structure ideal for semantic segmentation
- Preserves spatial context through skip connections

LSTM

Detects progressive temporal deterioration



U-Net

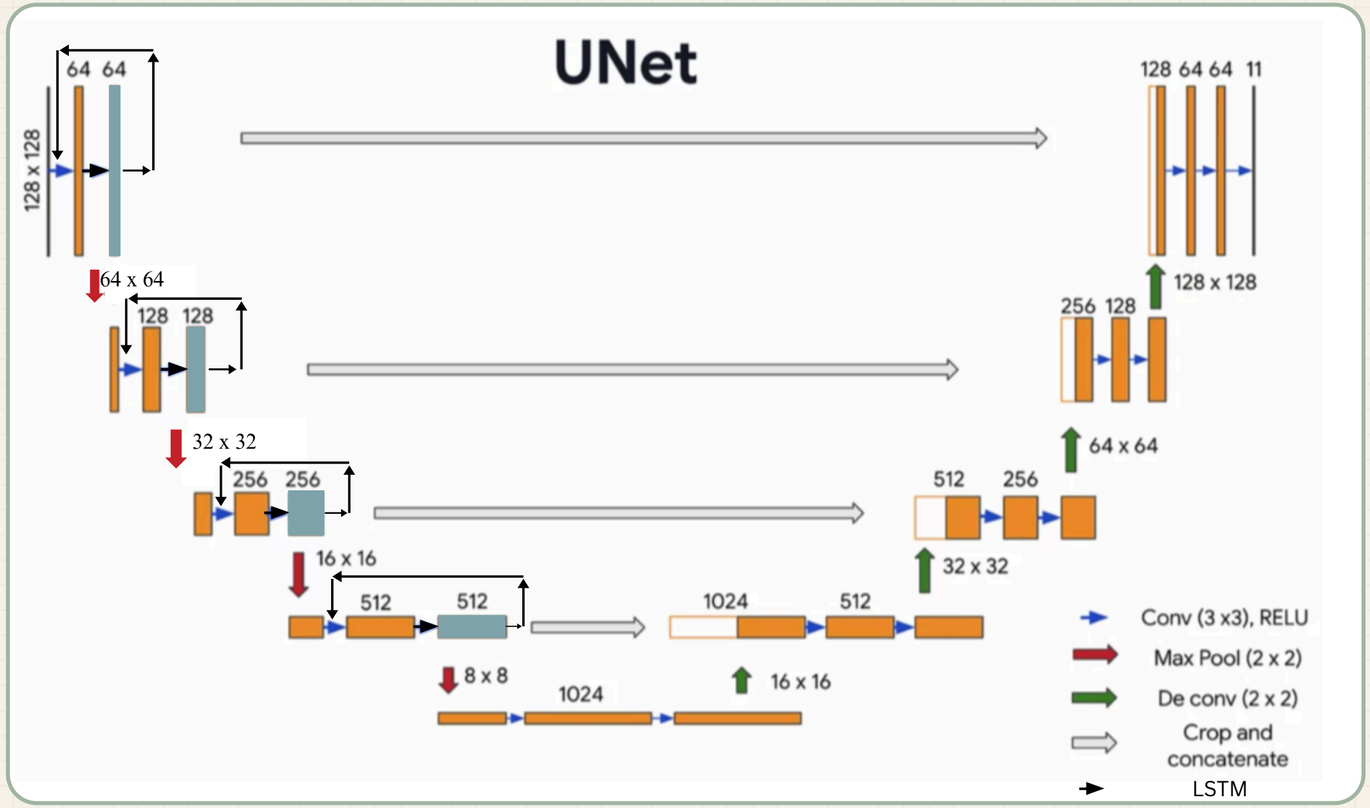
MODEL ARCHITECTURE

Encoder: Convolutional + ReLU + MaxPooling layers

convLSTM processes temporal sequence

Decoder: Upconvolution + concatenation with encoder features

Output: Pixel-level segmentation map with multiple damage classes



U-Net + convLSTM

TRAINING AND OPTIMIZATION

Loss Function

- **Dice Loss** to handle class imbalance for multiclass segmentation
- ε is “smooth” (10e-6) which ensures that loss is never divided by zero

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_{pred_i} y_{true_i} + \epsilon}{\sum_{i=1}^{N} y_{pred_i} + \sum_{i=1}^{N} y_{true_i} + \epsilon}$$

Dice Loss Function

Optimizer: Adam optimizer

- Learning rate: 10e-3
- Batch size: 2
- Epochs : 100

Regularization

Dropout layers, early stopping

Server/Tools

Kaggle-GPU P100

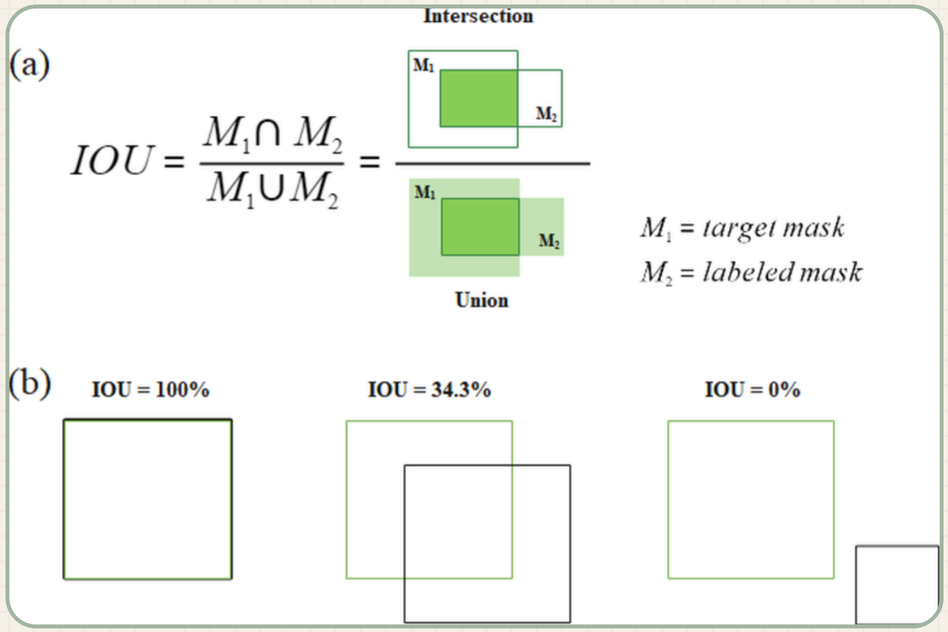
EVALUATION METRICS

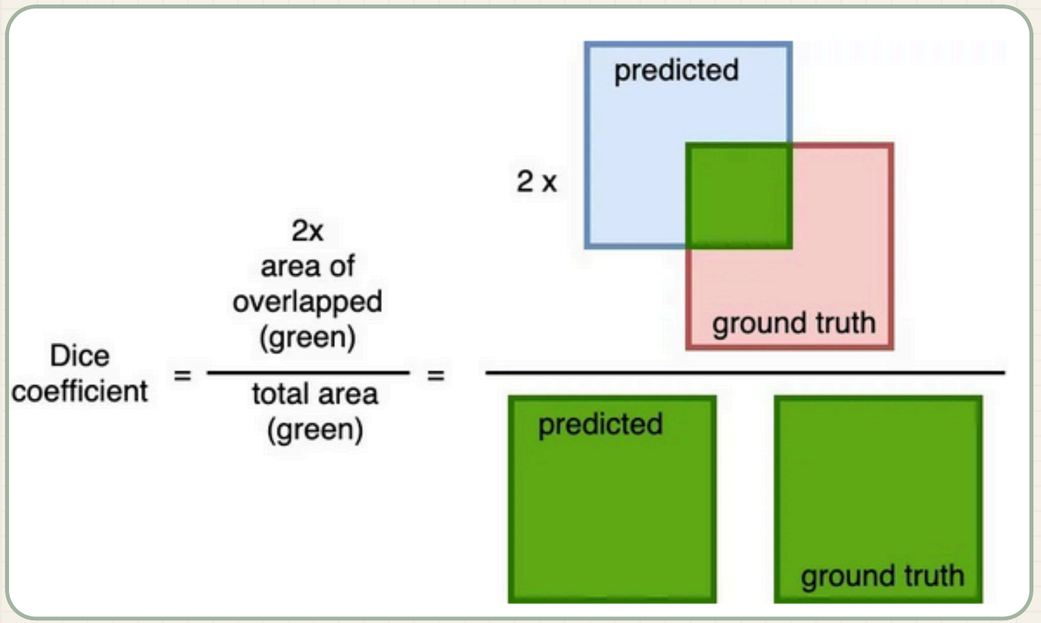
Performance Metrics Used:

- Mean Intersection over Union (IoU)
- Dice Coefficient

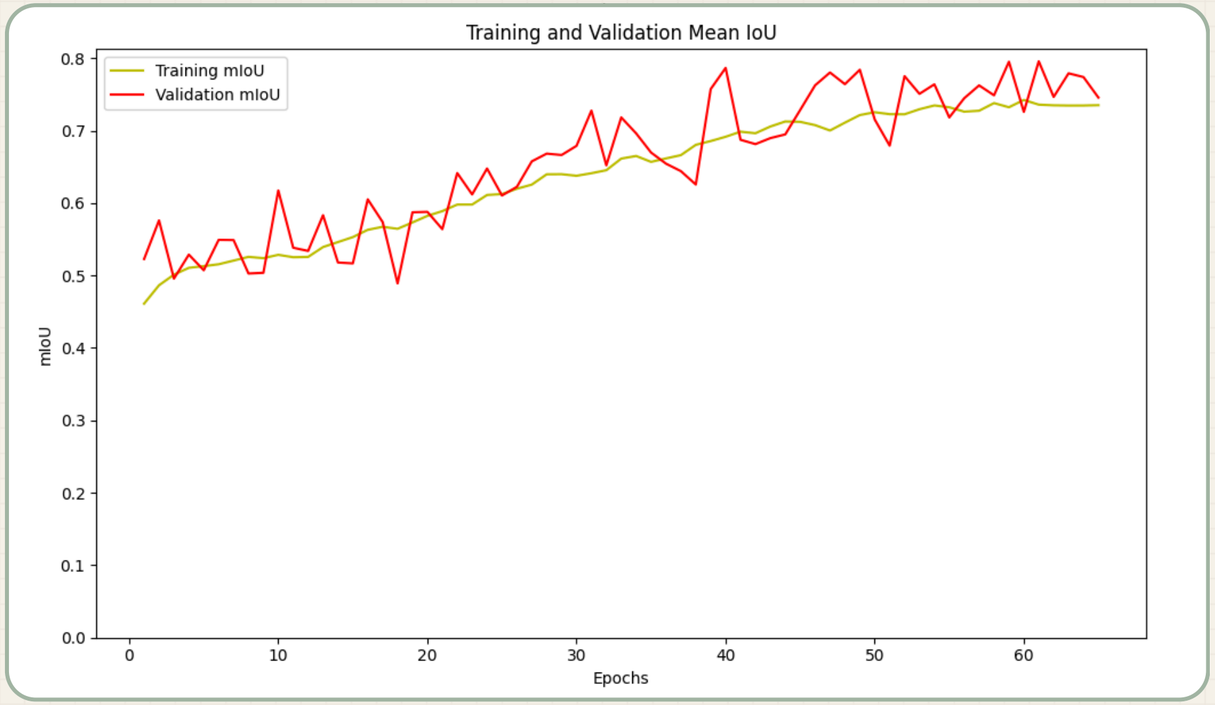
Qualitative Evaluation

- Visual comparison between predicted masks and ground truth





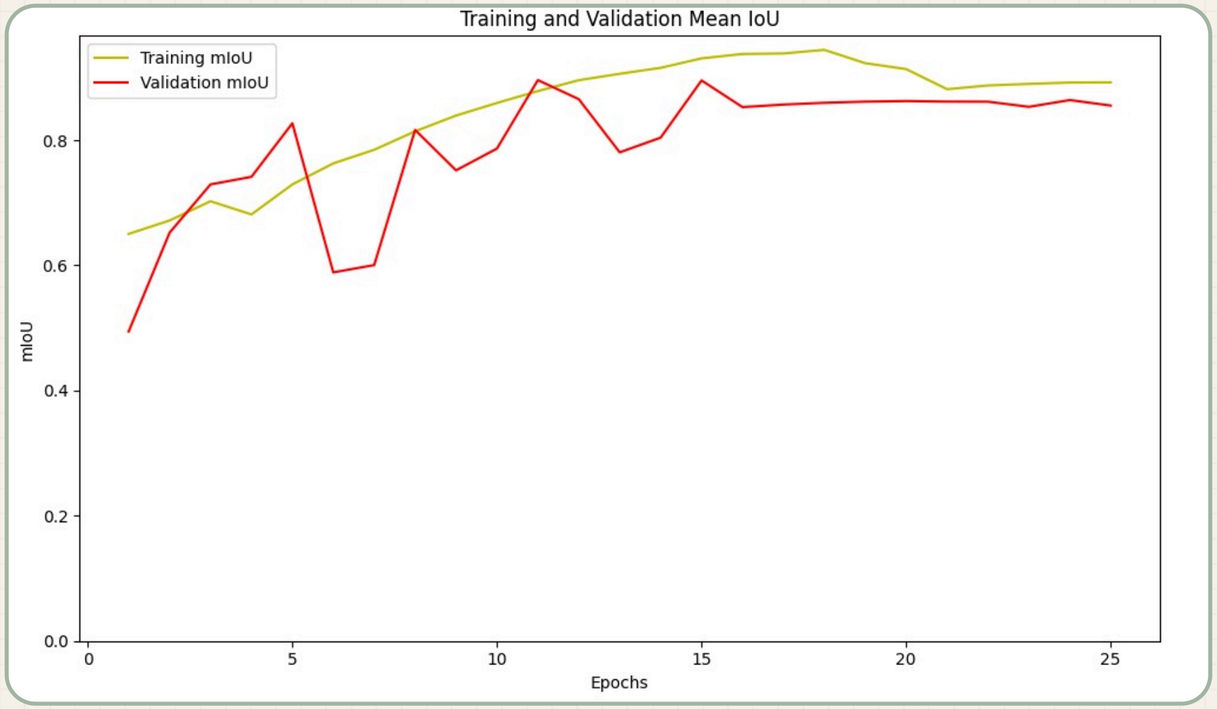
MODEL PERFORMANCE-U-NET



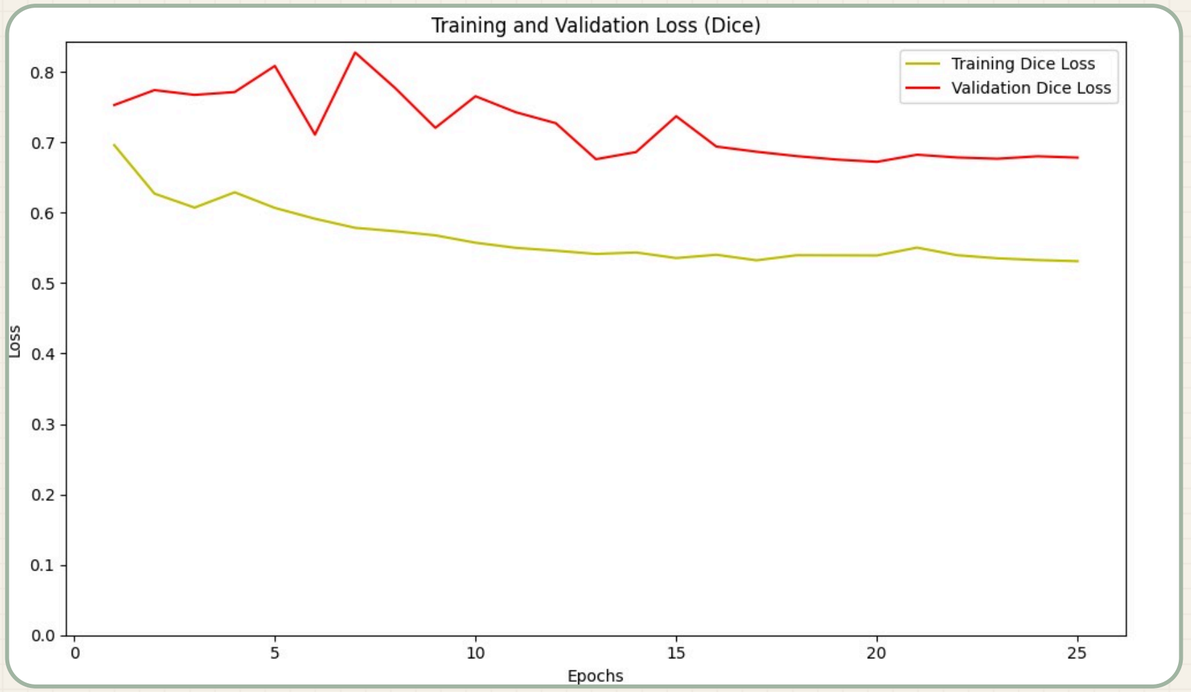
mIoU vs. Epochs

Validation Loss vs Epochs

MODEL PERFORMANCE-U-NET+CONVLSTM



mIoU vs. Epochs



Training vs Validation Loss

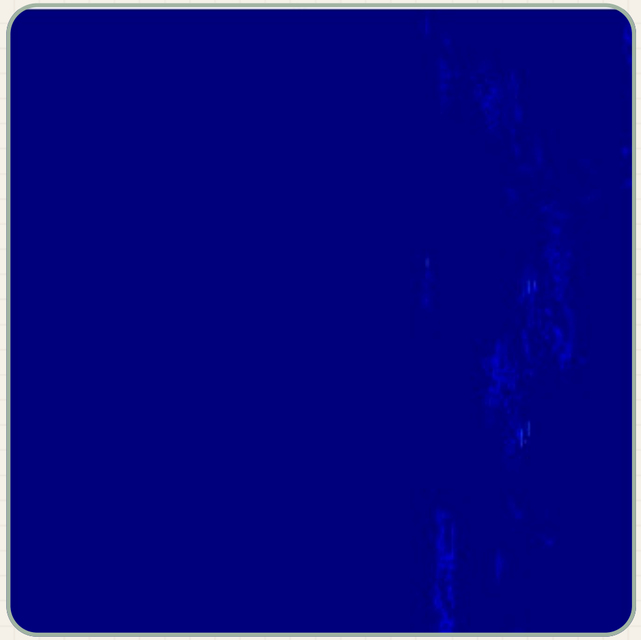
MODEL PERFORMANCE

Result Comparison for Test Dataset

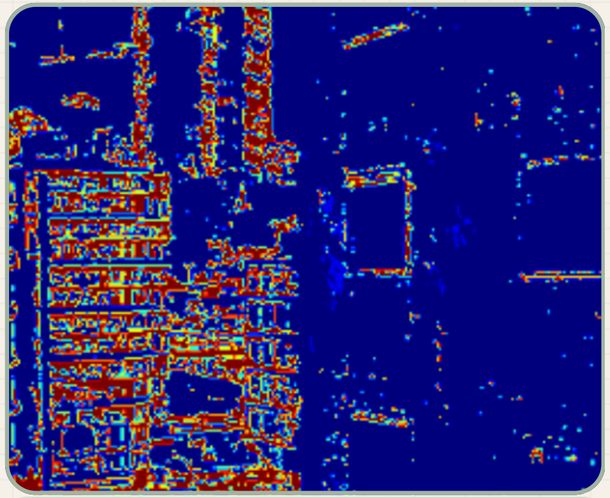
Metric	U-Net Only	U-Net + LSTM
Loss	0.56	0.55
Mean IoU	0.80	0.89
Dice Coefficient	0.43	0.45

MODEL PERFORMANCE

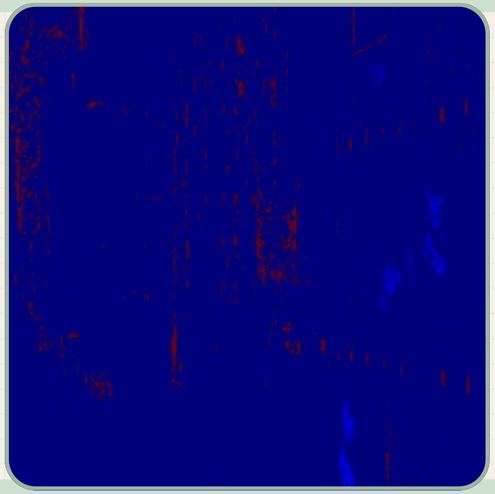
Original Image



Ground Truth Mask



Predicted Mask (U-Net)



Predicted Mask (U-Net + convLSTM)

Comparison of predicted segmentation maps — U-Net vs. U-Net + LSTM

DISCUSSION

- The synthetic dataset successfull trained deep learning models to assess structural conditions.
- U-Net with ConvLSTM was used to capture features better
- The ConvLSTM2D improves the detection of small, complex features by giving the network "**local memory**" to filter out noise and emphasize feature relationships across the image area.

Limitations:

Synthetic-to-Real Gap

The model trained on synthetic images may not work as well on real structures.

Computation

Adding LSTM makes training slower and needs more GPU power as no. of trainable param. increased to **40+ million**.

Multi Class Data

This priority-based method loses information by forcing pixels that belong to multiple classes.

Masks

Real lighting changes or blocked views aren't fully covered in the dataset

CONCLUSION

- Developed a vision-based framework for multiclass structural condition assessment.
- Utilized a large-scale synthetic dataset simulating realistic urban damage patterns.
- Tried Combining U-Net (spatial segmentation) with ConvLSTM
- Demonstrated the potential of synthetic data in training deep SHM systems.

REFERENCES

- https://sail.cive.uh.edu/quakecity/
- https://medium.com/@mhamdaan/multi-class-semantic-segmentation-with-u-net-pytorch-ee81a66bba89
- *Olaf Ronneberger, Philipp Fischer, Thomas Brox* 'U-Net: Convolutional Networks for Biomedical Image Segmentation'

Thankyou!
Any Questions?