Deep Learning-Based Crack Detection System Using Transfer Learning

Abstract

This report presents the development and evaluation of an automated crack detection system using deep learning techniques. The system achieves 99.81% accuracy in binary classification of cracked versus non-cracked surfaces using a modified ResNet18 architecture with transfer learning. The model demonstrates exceptional performance across all evaluation metrics, including 99.65% precision, 99.98% recall, and 99.81% F1-score. Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) visualization is implemented to provide interpretable crack localization without requiring manually annotated bounding boxes. The system includes real-time webcam detection capabilities for practical deployment scenarios.

---

1. Introduction

1.1 Background

Crack detection in structures and surfaces is critical for maintenance, safety assessment, and quality control across various industries. Traditional manual inspection methods are time-consuming, subjective, and prone to human error. Deep learning-based computer vision offers an automated, objective, and scalable solution for crack detection.

1.2 Objectives

The primary objectives of this project were to:

1. Develop a binary classification model to distinguish between cracked and non-cracked images
2. Achieve high accuracy (>95%) suitable for practical applications
3. Implement crack localization using explainable AI techniques
4. Create a real-time detection system using webcam input
5. Evaluate model performance using industry-standard metrics

1.3 Scope

This project focuses on binary image classification rather than pixel-level crack segmentation, making it suitable for scenarios where only crack/no-crack labels are available, avoiding the need for expensive bounding box annotations.

---

2. Methodology

2.1 Dataset

Dataset Composition:

- Total Images: 40,000
  - Positive samples (with cracks): 20,000 images
  - Negative samples (without cracks): 20,000 images
- Train-Validation Split: 80-20 ratio
  - Training set: 32,000 images (16,000 per class)
  - Validation set: 8,000 images (4,000 per class)

Data Preprocessing:

- All images resized to 224×224 pixels (standard ResNet input size)
- RGB color format maintained
- Balanced dataset to prevent class bias

Data Augmentation (Training Set Only):

- Random horizontal flipping
- Random rotation (±15 degrees)
- Color jittering (brightness and contrast variation ±20%)
- Normalization using ImageNet statistics (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])

2.2 Model Architecture

Base Model: ResNet18 (Residual Network with 18 layers)

Transfer Learning Approach:

- Pre-trained weights from ImageNet (1.2M images, 1000 classes)
- Modified final fully connected layer for binary classification
- Architecture modification:
- `text`

```
Original FC: 512 → 1000 classes
Modified FC: 512 → Dropout(0.3) → 2 classes
```

-

Model Parameters:

- Total trainable parameters: ~11.2 million
- Input size: 224×224×3
- Output: 2 classes (crack, no_crack)

Rationale for ResNet18:

- Proven effectiveness in crack detection research achieving 99%+ accuracy
- Fast inference time (suitable for real-time applications)
- Residual connections prevent vanishing gradient problem
- Good balance between accuracy and computational efficiency

2.3 Training Configuration

Hardware:

- GPU: NVIDIA GPU with CUDA support
- Framework: PyTorch 2.9.0

Hyperparameters:

- Optimizer: Adam (Adaptive Moment Estimation)
- Learning rate: 0.0001
- Batch size: 32
- Number of epochs: 5
- Loss function: Cross-Entropy Loss
- Learning rate scheduler: StepLR (decay by 0.1 every 7 epochs)

Training Strategy:

- Transfer learning with fine-tuning (all layers trainable)
- Early stopping implemented (saves best validation accuracy model)
- Training time: ~18 minutes total (5 epochs)

2.4 Explainable AI: Grad-CAM Visualization

To address the "black box" nature of deep learning models, Gradient-weighted Class Activation Mapping (Grad-CAM) was implemented to visualize which regions of the image the model focuses on when making predictions. This provides:

- Crack localization without manual bounding box annotations
- Model interpretability for validation and debugging
- Visual confidence assessment for predictions

---

3. Results

3.1 Training Performance

The model demonstrated rapid convergence and consistent improvement across all 5 epochs:

Epoch	Train Loss	Train Acc	Val Loss	Val Acc	Best Model
1/5	0.0149	99.51%	0.0072	99.80%	✓ Saved
2/5	0.0069	99.80%	0.0117	99.81%	✓ Saved
3/5	0.0050	99.81%	0.0068	99.78%	-
4/5	0.0050	99.85%	0.0064	99.75%	-
5/5	0.0046	99.87%	0.0068	99.80%	-

Key Observations:

- Best validation accuracy (99.81%) achieved at Epoch 2
- Minimal overfitting (train-val accuracy gap < 0.1%)
- Consistent loss reduction across epochs
- Training time per epoch: ~3.5 minutes (training) + 0.5 minutes (validation)

3.2 Final Model Performance Metrics

The model was evaluated on the validation set (8,000 images) using standard classification metrics:

Metric	Value	Definition
Accuracy	99.81%	Ratio of correct predictions to total predictions
Precision	99.65%	Of predicted cracks, % that are truly cracks (TP / [TP + FP])
Recall	99.98%	Of actual cracks, % detected by model (TP / [TP + FN])
F1-Score	99.81%	Harmonic mean of precision and recall (2 × P × R / [P + R])
Specificity	99.65%	Of actual non-cracks, % correctly identified (TN / [TN + FP])

Confusion Matrix Analysis:

- True Positives (TP): ~3,999 (cracks correctly identified)
- True Negatives (TN): ~3,986 (non-cracks correctly identified)
- False Positives (FP): ~14 (non-cracks misclassified as cracks)
- False Negatives (FN): ~1 (cracks missed)

3.3 Performance Interpretation

Exceptional Recall (99.98%): The model successfully detects nearly all actual cracks, critical for safety applications where missing a crack could have serious consequences.

High Precision (99.65%): Very few false alarms, making the system practical for deployment without overwhelming users with false positives.

Balanced Performance: The near-equal precision and specificity (both 99.65%) indicate the model performs equally well on both classes, demonstrating no class bias despite starting with a balanced dataset.

F1-Score (99.81%): The harmonic mean confirms the model achieves an optimal balance between precision and recall, making it suitable for production deployment.

3.4 Comparison with Literature

Research in crack detection using deep learning reports varying results:

Study	Model	Dataset Size	Accuracy	F1-Score
Golding et al.	VGG16	40,000	-	99.33-99.55%
Li et al.	AlexNet	6,000	99.06%	-
Our Study	ResNet18	40,000	99.81%	99.81%

Our model achieves state-of-the-art performance comparable to or exceeding existing literature, particularly impressive given only 5 training epochs versus 10-20 epochs in comparable studies.

---

4. Discussion

4.1 Transfer Learning Effectiveness

The exceptional results with only 5 epochs demonstrate the power of transfer learning. The pre-trained ResNet18 model already learned robust feature representations from ImageNet, requiring minimal fine-tuning for crack detection. This is particularly valuable for:

- Limited training time scenarios (10-day internship)
- Reduced computational costs
- Faster iteration during development

4.2 Data Quality Impact

The balanced dataset (50-50 crack/no-crack) prevented class imbalance issues that commonly plague binary classification tasks. The 40,000-image dataset size aligns with research showing that larger, diverse datasets significantly improve model generalization.

4.3 Grad-CAM for Interpretability

The implementation of Grad-CAM addresses a critical limitation in deploying AI systems in safety-critical applications. By visualizing which image regions influence predictions, the system provides:

- Verification that the model focuses on actual crack regions (not artifacts)
- Debugging capability when predictions seem incorrect
- Trust-building for end users who can see "why" the model made its decision
- Weak supervision for crack localization without expensive manual annotations

4.4 Practical Deployment

The real-time webcam detection feature demonstrates the system's readiness for practical deployment scenarios:

- Frame capture and analysis in <1 second
- Visual feedback with confidence scores
- User-friendly interface requiring no technical expertise

4.5 Limitations

Dataset Specificity: Model performance may vary on crack types not represented in training data (different materials, lighting conditions, crack patterns).

Binary Classification Only: The system identifies presence/absence of cracks but does not measure crack severity, width, or length.

Grad-CAM Approximation: While Grad-CAM highlights relevant regions, it provides an approximation rather than precise pixel-level crack boundaries.

Computational Requirements: GPU acceleration is recommended for training and beneficial for real-time inference.

---

5. Conclusion

This project successfully developed a highly accurate crack detection system achieving 99.81% accuracy using transfer learning with ResNet18. The model demonstrates:

✅ State-of-the-art performance with 99.81% F1-score
✅ Near-perfect recall (99.98%) for safety-critical applications
✅ High precision (99.65%) minimizing false alarms
✅ Fast training (18 minutes for 5 epochs)
✅ Explainable predictions via Grad-CAM visualization
✅ Real-time capability with webcam integration