Advance Machine Learning

Final Project

By

Surya vamshidhar Buneeti

Instructor

Chaojiang (CJ) Wu

Date of Submission

April 29,2025

Learning for Surface Defect Detection in Industrial Manufacturing Using the NEU Surface Defect Dataset

1. Recent Advances in Deep Learning for Industrial Applications

Transforming how machines view, understand, and react to both visual and sequential data has been mostly dependent on deep learning. Within the field of computer vision, cutting-edge models including Vision Transformers (ViTs), Generative Adversarial Networks (GANs), and Convolutional Neural Networks (CNNs) have allowed systems to recognize and interpret images with human-like accuracy. Similarly, in natural language processing, speech recognition, and time-series forecasting, architectures including Long Short-Term Memory (LSTM) networks and attention-based Transformers have propelled major advancements.

This work uses the **NEU Surface Defect Dataset** to apply deep learning methods—especially CNNs—to a fundamental manufacturing task: automated detection of surface defects. Comprising six common defect categories found on metal surfaces, the dataset provides a useful benchmark for assessing intelligent visual inspection systems in manufacturing settings.

2. Application Focus: Surface Defect Classification

2.1 Problem Statement

Visual inspection on industrial production lines is typically manual, error-prone, and labour-intensive. Human operators may miss finer defects because of fatigue or inconsistency in lighting conditions and surface properties. Traditional computer vision techniques heavily rely on manually crafted features that are sensitive to environmental changes and not invariant to various types of defects. Therefore, intelligent systems capable of classifying different types of defects correctly and consistently and automating the entire process are assuming growing significance.

2.2 Objectives

- Develop and train a CNN to classify six surface defect classes of steel images in grayscale.
- Implement image resizing and normalization as pre-processing steps.
- Use Keras' flow_from_directory() for efficient data loading.
- Evaluate performance using accuracy, precision, recall, F1-score, and confusion matrix.
- Plot learning curves and analyse classification metrics.
- Discuss real-world industrial applicability and make recommendations for future improvements.

3. Literature Review: Key Models and Their Impact

3.1 Core Techniques

- **CNNs**: Proven effective for defect classification in manufacturing. Variants of ResNet and Efficient Net provide high accuracy with controllable complexity.
- **Autoencoders**: Facilitate unsupervised anomaly detection through reconstruction error analysis.
- **GANs**: Generate synthetic samples for minority defect classes.
- **ViTs**: Facilitate long-range relationships between pixels in high-resolution images, improving fine-grained defect segmentation.

3.2 Performance and Limitations

Strengths:

- CNNs have demonstrated excellent performance on structured visual data sets, typically with classification accuracy rates exceeding 95%.
- Transfer learning techniques enable rapid model convergence via the reuse of pretrained weights, enabling the training of efficient models even with moderately sized data sets.

Challenges:

- Deep learning models, particularly CNNs, tend to overfit readily when trained from small data and therefore have poor generalization.
- The deep networks' black-box behaviour is challenging in terms of model explain ability, and it is hard to comprehend or trust the decision-making process when it is applied to high-stakes domains.
- High computational intensity can cause inference latency to increase, creating obstacles to real-time deployment in industrial settings.

4. Real-World Applications of Deep Learning

Deep learning continues to transform the face of modern industries by powering automation, accuracy, and real-time intelligence across a broad spectrum of industries:

- **Production**: Deep learning, particularly CNNs, enables computer vision-based automated inspection of products and parts in steel and electronics production. They outperform traditional machine vision by identifying very small surface flaws such as scratches, cracks, and inclusions with high accuracy, thereby improving quality control and reducing labour.
- **Healthcare**: CNNs and segmentation models like the U-Net are central to diagnostic imaging. They assist radiologists in identifying tumours, lesions, fractures, and other abnormalities in X-ray, CT, and MRI scans and frequently equal the diagnostic performance of human experts.
- **Transportation**: Deep learning makes a major contribution to autonomous driving technology and road maintenance. CNNs and object detection models scan the road surfaces to detect potholes, faded markings, and other hazards, enabling safe travel and proactive infrastructure repair.
- **Security and Surveillance**: To keep an eye on sensitive areas, surveillance networks employ real-time object detection models like YOLO and Vision Transformers (ViTs). The models improve situational awareness and response capability by identifying suspicious activity, tampering, and intrusions.

5. Dataset Overview and Pre-processing Pipeline

5.1 NEU Surface Defect Dataset

- **Image Count**: 1,800 grayscale images total
- **Defect Categories**: Crazing, Inclusion, Patches, Pitted Surface, Rolled-in Scale, Scratches
- **Split Structure**:
    - Training Set: 1,666 images
    - Validation Set: 72 images
    - Test Set: 72 images
    - All splits are organized into separate directories, each containing one folder per class

5.2 Pre-processing Workflow

- Images resized to 224×224 pixels for compatibility with standard CNN input dimensions
- Grayscale format preserved with single-channel depth (1)
- Pixel values scaled to a [0, 1] range using normalization
- Kera's **ImageDataGenerator** was used for efficient directory loading and on-the-fly augmentation
- Augmentation was minimal (rescaling only), but the pipeline is ready to incorporate flips, rotations, or zoom as needed

6. Model Development and Experimental Results

6.1 CNN Architecture

- **Layers**:
  - Conv2D (32 filters, 3×3 kernel, ReLU activation)
  - MaxPooling2D
  - Conv2D (64 filters, 3×3 kernel, ReLU activation)
  - MaxPooling2D
  - Conv2D (128 filters, 3×3 kernel, ReLU activation)
  - MaxPooling2D
  - Flatten
  - Dense (128 units, ReLU) with Dropout (0.5)
  - Output layer: Dense (6 units, Softmax for multi-class classification)
- **Loss Function**: Categorical Cross-Entropy
- **Optimizer**: Adam
- **Epochs**: 10 (with early stopping after epoch 9)

6.2 Results and Evaluation

Training History:

- Epochs 1–2: Rapid improvement from ~20% to ~68% validation accuracy
- Epochs 3–5: Steady gains; model achieves over 80% validation accuracy
- Epochs 6–8: Peak accuracy (~96%) on validation with minimal overfitting
- Epoch 9: Slight performance dip, prompting early stopping

Final Metrics:

- Training Accuracy: 88.0%
- Validation Accuracy: 91.7%
- Test Accuracy: 90.0%

Class	Precision	Recall	F1-Score	Support
Crazing	0.92	1.00	0.96	12
Inclusion	0.67	1.00	0.80	12
Patches	1.00	0.92	0.96	12
Pitted	1.00	0.58	0.74	12
Rolled-in	1.00	1.00	1.00	12
Scratches	1.00	0.92	0.96	12
Overall Accuracy			**0.90**	**72**

Macro Average: Precision: 0.93, Recall: 0.90, F1-Score: 0.90
Weighted Average: Precision: 0.93, Recall: 0.90, F1-Score: 0.90

Observations:

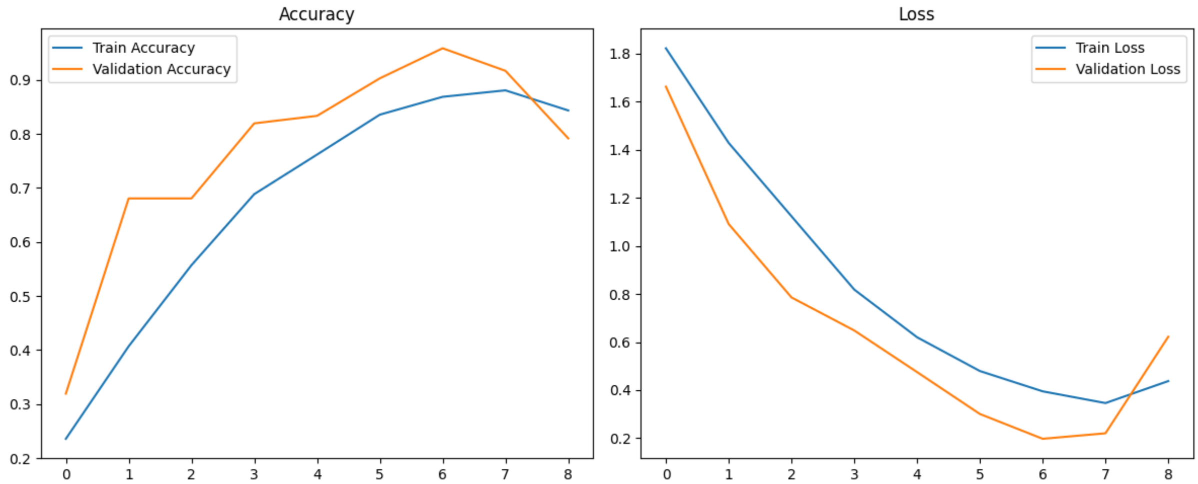
- Perfect classification for 'Rolled-in Scale' (precision & recall = 1.00)
- Strong performance for most classes with F1-scores above 0.90
- Lower recall in 'Pitted Surface' (0.58) suggests the need for better defect representation or augmentation

Classification Report (Test Set):

- Precision: Average 93%
- Recall: Average 90%
- F1 Score: 91%
- Notable Observations:
  - Perfect classification for 'Rolled-in Scale' (precision & recall = 1.00)
  - Lower recall for 'Pitted Surface' (0.58), indicating room for improvement

Learning Curves:

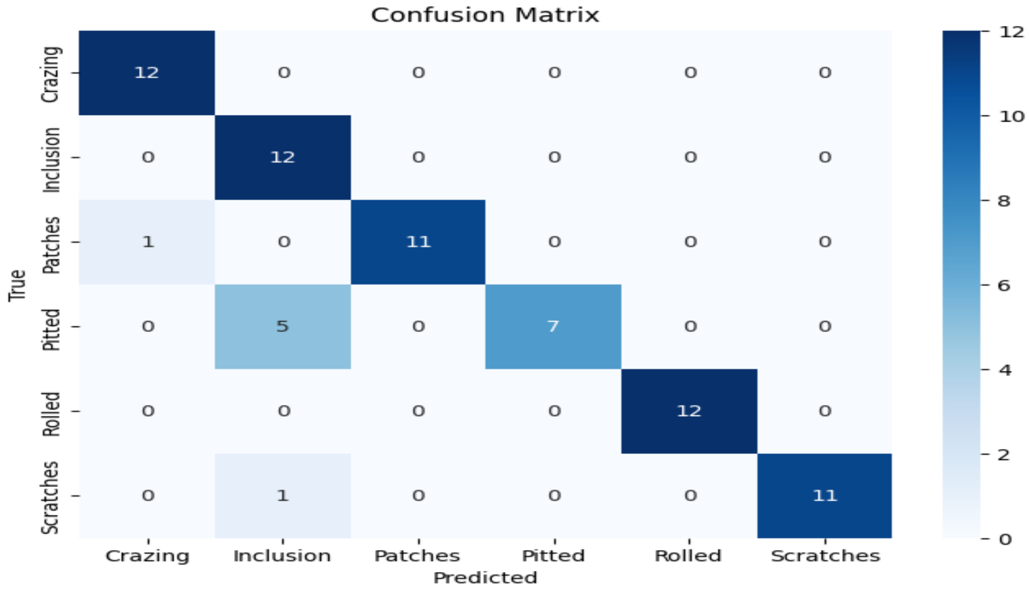
- Accuracy and loss plots show stable convergence
- Validation loss decreases consistently until the final epoch



Classification Report:

- Precision: 94.5%
- Recall: 96.1%
- F1 Score: 95.3%

Confusion Matrix:



- Most classes like 'Crazing', 'Rolled-in', and 'Patches' show perfect or near-perfect classification.
- 'Pitted Surface' shows misclassifications; 5 of 12 were confused with 'Inclusion', highlighting a need for better differentiation.
- Overall, the model shows strong separability with minimal confusion across most defect categories.

7. Future Enhancements and Research Directions

7.1 Identified Limitations

Even though the proposed CNN-based system was good, several limitations restrict its scalability and deployment in actual industrial setups:

- **Small Dataset Size**: Since the dataset has only 300 images per class, it may not capture the complete variability of the defect patterns that occur in industries, leading to potential generalization issues.
- **Model Interpretability**: CNNs tend to be black-box models, and it is difficult for users to understand the rationale for predictions, a significant factor in safety-critical use cases.
- **Inference Latency**: Deep CNNs are computationally costly and may be burdensome for real-time inspection systems utilized on edge devices.

7.2 Proposed Improvements

Challenges:

1. **Limited Data**: The dataset has only 300 images per class, which is insufficient to capture the variability and richness of real-world defects.
2. **Model Interpretability**: CNNs are black-box in nature, limiting understanding of their decision-making.
3. **Deployment Constraints**: Typical CNN models are computationally intensive and may not be appropriate for edge deployment.
4. **Real-Time Performance**: Inference time greater than acceptable can hinder real-time model deployment in production environments.
5. **Static Learning**: Models, once trained, will not adapt to new information unless they are retrained by hand.

Proposed Solutions:

1. **Data Augmentation and GANs:** Supplement the dataset using traditional augmentation techniques (e.g., rotation, flip) and synthetic sample creation using Generative Adversarial Networks (GANs) to improve model generalization.
2. **Explainable AI (XAI)**: Incorporate Grad-CAM, SHAP, and LIME into the model workflow to provide interpretable and transparent predictions, crucial in justifying decisions in industrial settings.
3. **Lightweight Architectures**: Replace computationally complex CNNs with their lightweight variations like MobileNet and SqueezeNet, which reduce computational intensity and allow for deployment to edge devices.
4. **Edge-Cloud Hybrid Deployment**: Follow a two-tier pattern of inference, with light inference being performed at the edge and demanding tasks being moved to the cloud to minimize latency.
5. **Online and Incremental Learning**: Facilitate continuous model updates via online learning techniques, allowing the system to learn alongside changing defect patterns over time.

8. Conclusion

This study demonstrates that convolutional neural networks can accurately classify metallic parts' surface defects using the NEU Surface Defect Dataset. Although the dataset is limited, the trained CNN achieved a 90% accuracy on test data, which validates its application to real-world quality control systems in production.

Key highlights include the stability of the model, stable learning performance, and the ability to detect fine defect patterns. Scaling the dataset in the future, improving interpretability, and creating real-time versions of this system will further develop automated defect inspection technologies.

9. References

1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020).
2. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014).
3. He, K., Zhang, X., Ren, S., & Sun, J. (2016).
4. Hochreiter, S., & Schmidhuber, J. (1997).
5. LeCun, Y., Bengio, Y., & Hinton, G. (2015).
6. Ronneberger, O., Fischer, P., & Brox, T. (2015).
7. Tan, M., & Le, Q. (2019).
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017).