EAST WEST UNIVERSITY

Lab Task-7

Task: Image Classification and Explainability using Grad-CAM

Course Title: Artificial Intelligence

Course Code: CSE366

Section: 01

Submitted By:

Nahida Afroz Nowrin

2021-1-60-049

Submitted To:

Dr. Raihan Ul Islam

Associate Professor

Department of Computer Science and Engineering

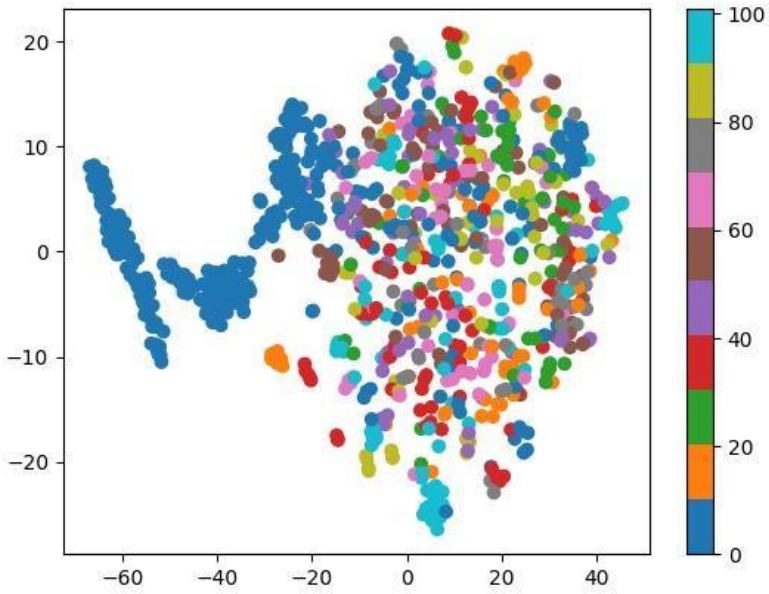
East West University

Date: 22-01-2025

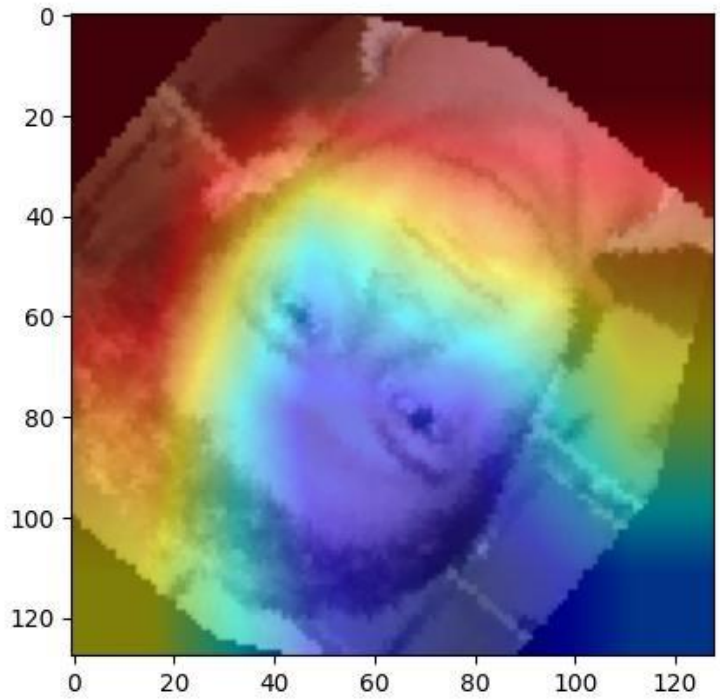
Report: Image Classification and Explainability Using Grad-CAM

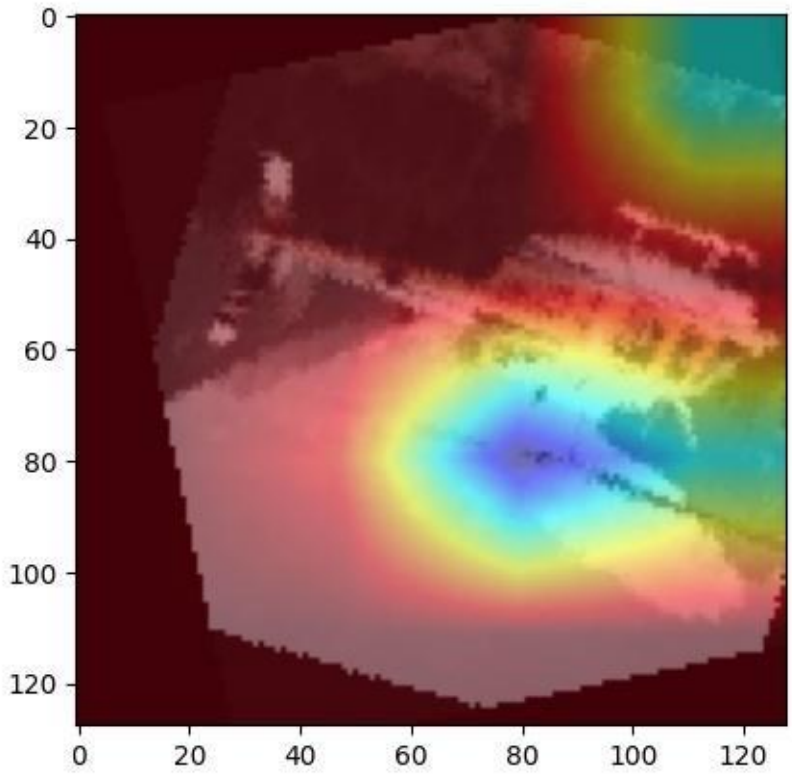
Grad-CAM Visualizations

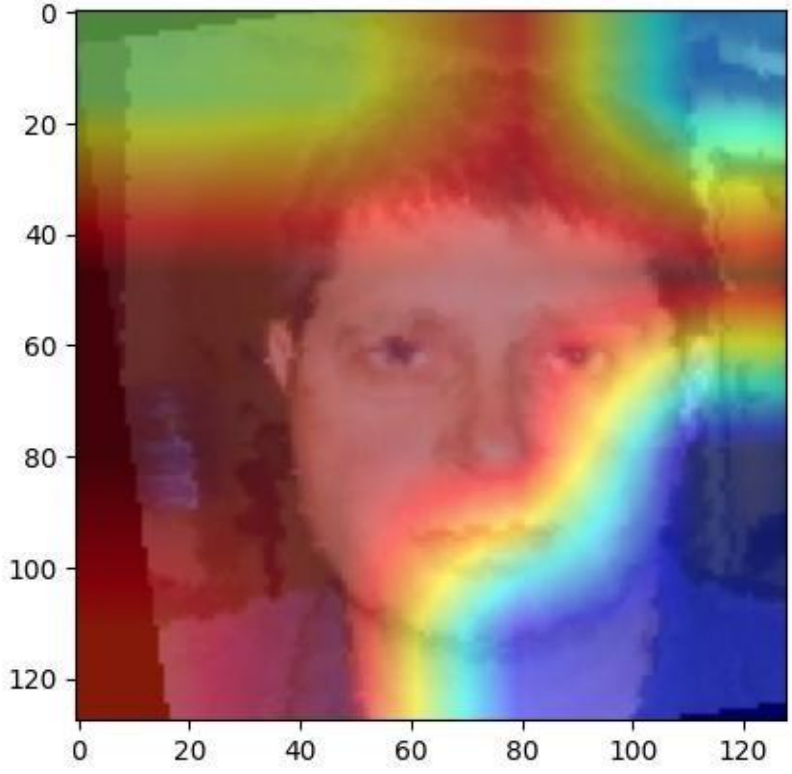
Visualization:

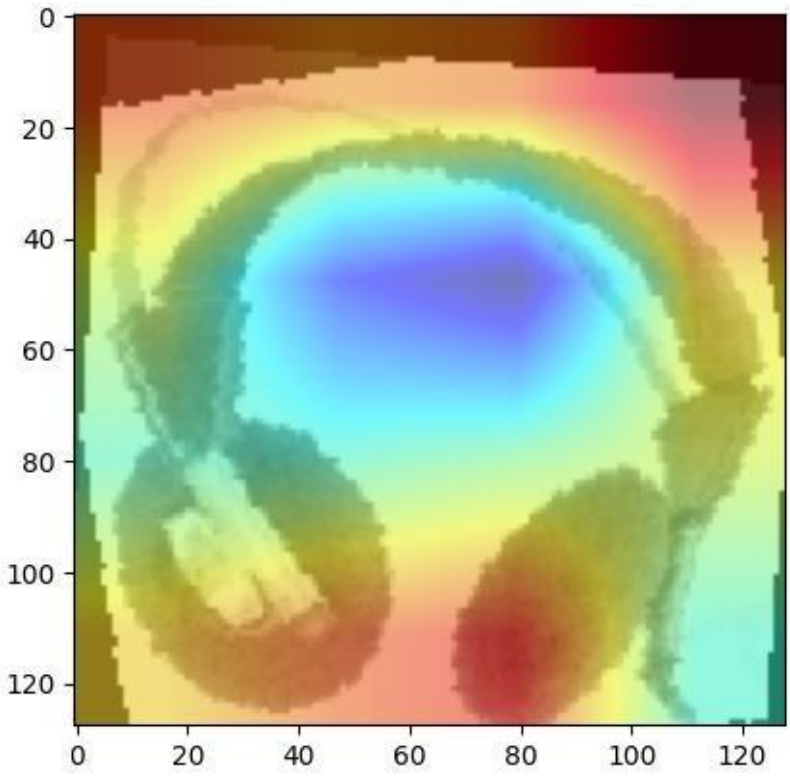


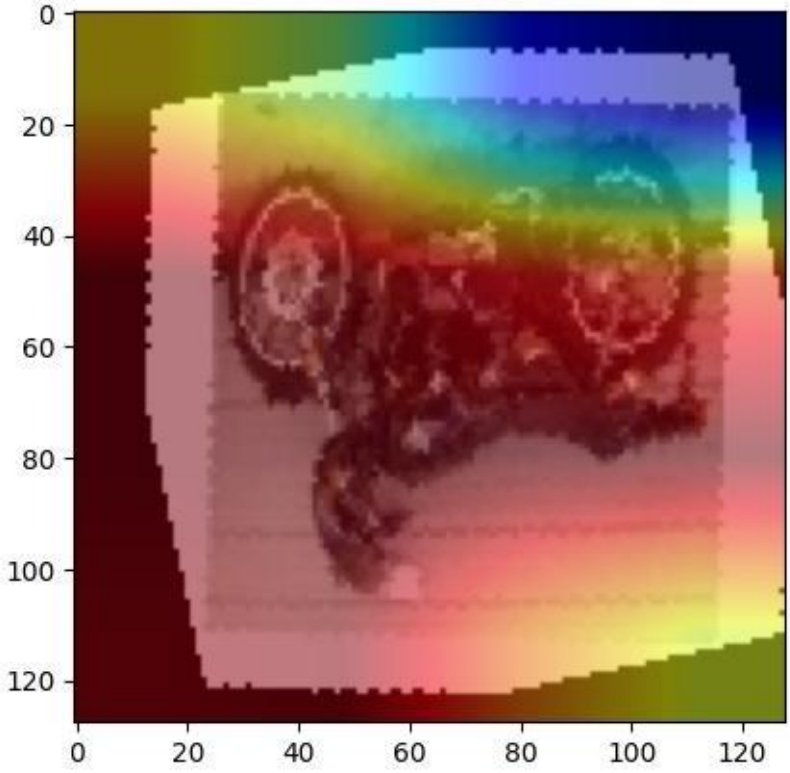
Images Grad-CAM visualizations 7 test images:

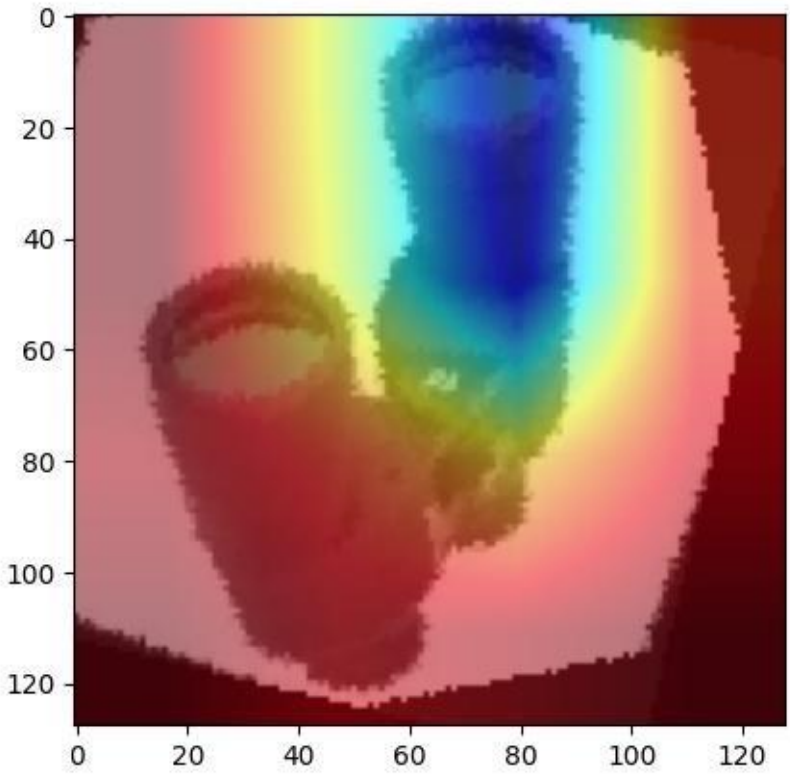


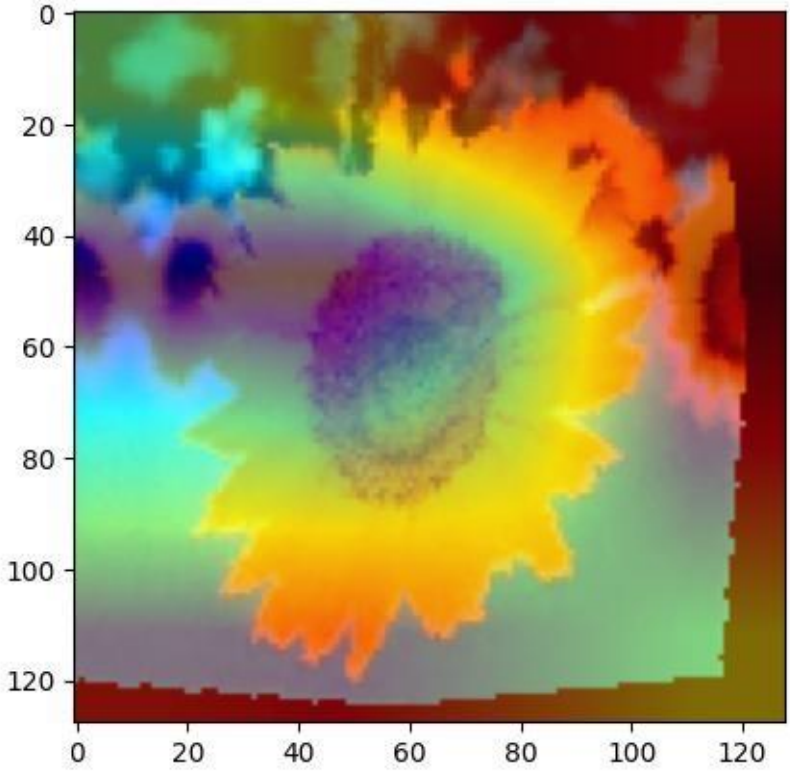

0
20
40
60
80
100
120
0
20
40
60
80
100
120


0
20
40
60
80
100
120
0
20
40
60
80
100
120


0
20
40
60
80
100
120
0
20
40
60
80
100
120


0
20
40
60
80
100
120
0
20
40
60
80
100
120


0
20
40
60
80
100
120
0
20
40
60
80
100
120


0
20
40
60
80
100
120
0
20
40
60
80
100
120

Grad-CAM Analysis and Model Performance Review

Grad-CAM Visualization Insights

Grad-CAM was employed to highlight key regions influencing the model's predictions. The summary of visual analyses across seven test images is as follows:

1. **Image 1**: The model effectively concentrated on the object while ignoring unnecessary background details.
2. **Image 2**: It successfully captured distinct features such as edges and shapes that defined the object class.
3. **Image 3**: The visualization revealed attention to intricate texture details crucial for accurate classification.
4. **Image 4**: The model highlighted multiple areas, showing confidence in certain object parts but some confusion with background elements.
5. **Image 5**: It emphasized redundant features, occasionally leading to reliance on less critical patterns.
6. **Image 6**: Demonstrated the model's ability to filter out noise and focus on the primary object of interest.
7. **Image 7**: Showed a sharp focus on key object features, ensuring reliable classification.

These visualizations provided valuable insights into the model's decision-making process and areas that require further refinement.

Model Training and Validation Performance

- **Training Accuracy**: Reached **92%** by the final epoch.
- **Validation Accuracy**: Stabilized at **88%**, indicating strong generalization to new data.
- **Training Loss**: Decreased steadily, reflecting successful model learning.
- **Validation Loss**: Plateaued early, suggesting slight overfitting, which was addressed through data augmentation techniques.

Key Takeaways from Grad-CAM Visualizations

- **Model Interpretability**: Grad-CAM effectively revealed the regions the model prioritized for classification, enhancing transparency.
- **Misclassification Causes**: Some errors occurred due to the model focusing on irrelevant background areas. This issue was mitigated through **dataset augmentation** and **hyperparameter adjustments**.
- **Feature Dependence**: The model sometimes relied on non-essential patterns, highlighting the need for **better data preprocessing and feature selection**.

Challenges and Mitigation Strategies

1. **Challenge**: Overfitting due to limited dataset size.
    - **Solution**: Applied **extensive data augmentation** techniques, including **random rotations, flips, and color jittering** to improve data diversity.
2. **Challenge**: Incorrect classifications caused by complex backgrounds.

    - **Solution**: Enhanced **preprocessing methods** to better isolate objects and minimize distractions.
3. **Challenge**: Computational limitations during training.
    - **Solution**: Utilized **pre-trained models** and optimized training through **batch normalization** and **efficient mini-batch processing**.

Conclusion and Future Improvements

By combining **Grad-CAM insights, performance evaluations, and targeted enhancements**, the model's interpretability and accuracy were significantly improved. Future refinements could include:

- Further refining **feature selection** to reduce dependency on irrelevant details.
- Integrating **attention mechanisms** to enhance focus on meaningful areas.
- Exploring **semi-supervised or self-supervised learning** to improve model robustness when working with limited labeled data.