# Deep Neural Networks for Anomalous Activity Detection

Naomi Sellam, *REU Student, Massbay Community College*
Shahriar Rifat, *PhD Candidate, Northeastern University*
Francesco Restuccia, *PhD, ECE Department, Northeastern University*

MENTIS LABORATORY

Northeastern University College of Engineering

Northeastern University Michael B. Silevitch and Claire J. Duggan Center for STEM Education

## Abstract

Without constant human monitoring, surveillance cameras often record anomalous activity that is discovered too late for effective response or missed entirely. To address this, we developed deep learning computer vision algorithms for **automatic detection of anomalies in video frames using convolutional neural networks**.

Using the **UCF Crime Dataset**, we implemented and tested **ResNet-50** and **MobileNet** to evaluate their performance in **flagging and classifying criminal activity**. Preliminary results show strong performance using **ResNet-50**, highlighting its potential for **real-time deployment** in **safety-critical environments**.

Ultimately, this technology aims to continuously **monitor live footage, detect and classify anomalous activity, alert users,** and **store the relevant video**.

## Background

Over **1 billion surveillance cameras** in homes, businesses, and public spaces record vast amounts of video daily. **Anomalous events** such as trespassing, theft, violence, and medical emergencies are frequently captured but rarely discovered or addressed due to **limited human oversight**. Traditional manual monitoring is error-prone, labor-intensive, unscalable, and highly inefficient, especially given the rarity of anomalies relative to normal activity. Yet timely detection is crucial to fulfilling the promise of safety these systems offer. Advances in deep learning enable continuous, **automated video frame analysis**, supporting **faster detection** and responses in both public spaces and private homes.


*Fighting*


*Normal activity*

To evaluate model performance in realistic conditions, we used the **UCF Crime Dataset**, a large collection of real surveillance footage labeled with 13 types of anomalous events including **robbery, vandalism, arson, road accidents,** and **assault**.

## Experimental Setting

### Dataset: UCF Crime

**Extracted frames from every video in UCF Crime**

- # of videos: **1900**
- Dataset length: **128 hours**
- # of classes: **13**
  - Abuse
  - Burglary
  - Robbery
  - Stealing
  - Shooting
  - Arrest
  - Assault
  - Fighting
  - Arson
  - Explosion
  - Shoplifting
  - Road Accident
  - Vandalism
  - Normal Event
- Image size: **64 x 64**
- Train subset: **1,266,345 images**
- Test subset: **111,308 images**
- Total: **1.38m video frame images**

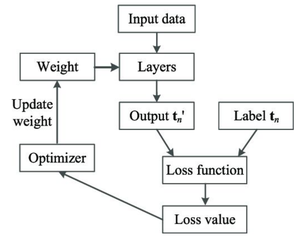

*Arson*


*Shoplifting*


*Road accident*


*Burglary*

### Training



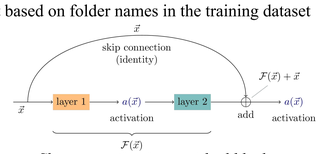

- Input images are fed through the model (e.g., ResNet-50 or MobileNet) to produce predictions.
- Predictions are compared to ground truth labels using a loss function (e.g., CrossEntropyLoss).
- The loss value quantifies prediction error.
- The optimizer (e.g., SGD) updates model weights to minimize the loss.
- Process repeats for each batch & epoch, only the model architecture differs.

### Architecture

#### ResNet-50

**Backbone**: Pretrained ResNet-50 imported from **torchvision.models**
**Feature Size**: Outputs a **2048**-dimensional feature vector from final convolutional block
**Architecture Type**: Uses standard convolutions with residual (skip) connections to enable deep feature learning
**Output Classes**: **num_classes** is set based on folder names in the training dataset
**Custom Classification Head**:

- **Linear(2048 → 512)**
- **ReLU**
- **Dropout(p=0.3)**
- **Linear(512 → num_classes)**




*Skip connection in a residual block*

Preserves important features and improves gradient flow, enabling deeper networks to train more effectively & accurately.

#### MobileNet

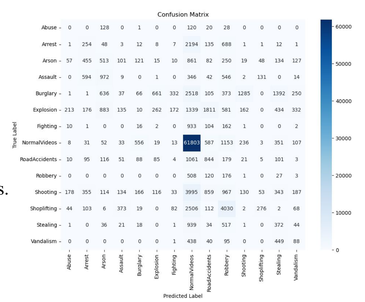
**Backbone**: Pretrained MobileNetV2 from **torchvision.models**
**Feature Size**: Outputs a 1280-dimensional feature vector from the final layer before classification
**Architecture Type**: Uses depthwise separable convolutions for efficient computation
**Inverted Residual Blocks**: Compact blocks with linear bottlenecks to reduce memory and computation
**Output Classes**: **num_classes** from folder names in the training dataset
**Custom Classification Head** (if matching your ResNet head setup):

- **Linear(1280 → 512)**
- **ReLU**
- **Dropout(p=0.3)**
- **Linear(512 → num_classes)**

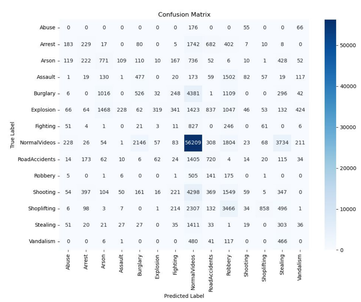

*Residual connection in MobileNetV2 bottleneck block*

## Results

- Correctly classified **61,803 NormalVideos**, indicating strong separation from anomalies.
- Significantly fewer false negatives:
  - Only **556 Assaults** and **7 Arrests** mislabeled as "NormalVideos".
- Produced **tighter confusion patterns** with fewer off-diagonal misclassifications.
- Heatmap shows **stronger diagonal intensity**, reflecting confident and accurate predictions.
- More robust for real-world anomaly detection, where missing critical events can be costly.




*Confusion matrix for ResNet-50*




*Confusion matrix for MobileNet*

- Correctly classified **56,209 NormalVideos**, showing weaker distinction from anomalies.
- Frequently mislabeled anomalies as normal:
  - **2,146 Assaults** and **1,742 Arrests** predicted as "NormalVideos".
- Showed more confusion between similar anomaly classes (e.g., Shoplifting vs. Stealing).
- Heatmap reveals **more off-diagonal errors**, indicating higher misclassification rates.
- Lower reliability for detecting rare or subtle anomalies in surveillance footage.

## Discussion

This project demonstrates that **modern CNN's can effectively classify suspicious activities in complex environments** if trained on a large corpus of data in real-life environments like those in UCF Crime Dataset. The preliminary results suggest a **pathway toward automated systems that augment human oversight** and reduce response time.

Deploying such technology in public and private spaces could **deter criminal behavior, assist law enforcement in timely intervention,** and **provide valuable evidence for post-event analysis**. However, it also raises important **ethical and privacy considerations**. Ensuring **transparency in model decisions,** minimizing **bias in training data,** and implementing **strict access controls** for stored footage are critical to maintaining public trust.

## Future Work

- **Continuous** monitoring capable anomaly detectors on **real-time** video frames
- **Segmentation and object detection** to identify suspicious persons or items related to the anomalous activity
- **Event-level understanding** using large language models (LLMs) for accurate description generation of anomalous events

## References

Real-World Anomaly Detection – CRCV. https://www.crcv.ucf.edu/projects/real-world/
UCF Crime Dataset – Kaggle. https://www.kaggle.com/datasets/odins0n/ucf-crime-dataset
Training Process of the CNN – ResearchGate https://www.researchgate.net/figure/Training-process-of-the-CNN_fig3_349987639
Skip Connection Diagram – TikZ.net. https://tikz.net/skip-connection/
MobileNet Architectures – Medium. https://medium.com/@pandrii000/mobilenet-architectures-17fe7406d794

## Acknowledgments

This project would not have been possible without your mentorship, insight, and unwavering support. Thank you.

Shahriar Rifat, PhD Candidate and Mentor
Northeastern University Center for STEM Staff
Brenda Egan, MBCC STEM Career Counselor and Mentor
MENTIS Lab Team
Francesco Restuccia, PhD, Principal Investigator

**NSF REU-Site: Research Experience for Undergraduates site to create "Pathways for community college students to enrich their education and careers" (REU-PATHWAYS)**

This work was supported by the National Science Foundation | Grant #2150417