ACCIDENT DETECTION USING DEEP NETWORKS

Saideep Arikontham
Pattern Recognition and Computer Vision - CS 5330
Data Science Graduate Student
Northeastern's Roux Institute, Portland, Maine

INTRODUCTION

This project develops and evaluates deep learning models (CNNs and ResNet) for live accident detection in CCTV footage. The goal is to create a system that accurately identifies accidents to improve surveillance, enable rapid response, and potentially save lives. The project focuses on optimizing model performance through various configurations and training/fine-tuning techniques, ultimately selecting the best model for live CCTV accident detection.

METHODOLOGY

Data collection

The project uses an image dataset from Kaggle containing training (369 accident, 422 non-accident images), validation (46 accident, 52 non-accident images), and test sets (47 accident, 53 non-accident images) for deep learning model development. These datasets will be used to train, validate, and select the best-performing model for live CCTV footage accident detection.

Data processing

Images were preprocessed by resizing them to 224x224 pixels, converting them to PyTorch tensors, and normalizing them using ImageNet statistics; additionally, 'Accident' class was labeled as 1 and 'Non Accident' as 0 for consistency across CNN and ResNet models. These steps ensure the images are in a suitable format for training the accident detection models.

Training CNNs

Initially, 50 random CNN architectures were trained, and the top 5 performers were selected based on validation accuracy. Optuna was then used to optimize the CNN architecture through hyperparameter tuning (varying convolutional/linear layers, dropout, hidden units, batch size), with models trained on preprocessed images and the best configuration saved after 40 trials.

METHODOLOGY

ResNet Training

ResNet18, ResNet34, and ResNet50 architectures were trained and evaluated, and ResNet18 was selected for fine-tuning based on its performance. The fine-tuning process used Optuna (for 20 trials) to optimize hyperparameters (dropout, frozen layers, learning rate, batch size, epochs) for a pretrained ResNet18 model, with the best model from each trial saved based on validation loss.

Model evaluation

After fine-tuning both the CNN and ResNet18 models, their performance was evaluated on a test set. The ResNet18 model demonstrated better generalization and was selected for live CCTV footage accident detection. Further analysis of the ResNet18 model's validation set performance at various classification thresholds determined that a threshold of 0.5 yields optimal results and will be used for live footage detection.

Live testing

A Kaggle-sourced video of combined CCTV and other accident footage was used to test the fine-tuned ResNet18 model. The video was processed by extracting frames, preprocessing them, and classifying each frame using the model with classification threshold 0.5. The output is a labeled video where each frame is captioned with the accident prediction based on those thresholds.

FINE-TUNING RESULTS

BEST CNN CONFIGURATION

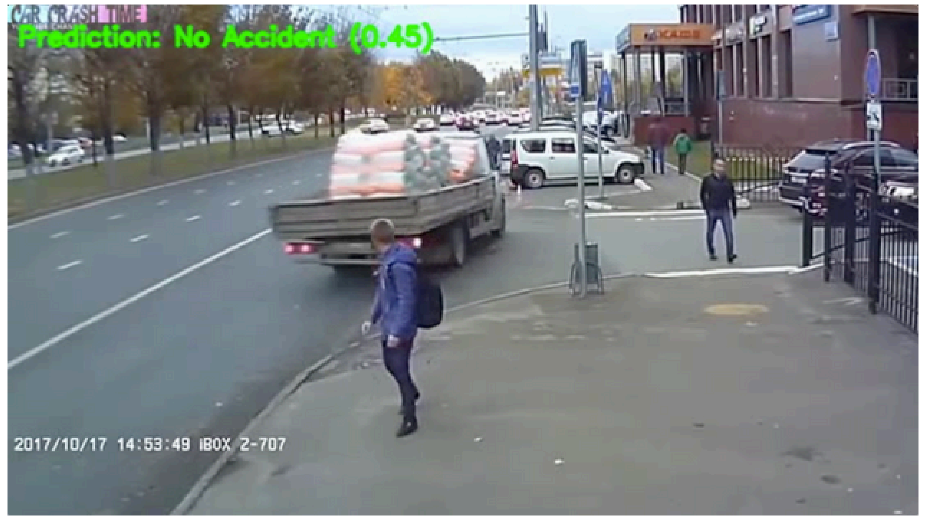
```
Trial 23 | Epoch 1/16 | Train Loss: 0.6718 | Val Loss: 0.6132 | Val Acc: 0.6633
Trial 23 | Epoch 2/16 | Train Loss: 0.6052 | Val Loss: 0.5735 | Val Acc: 0.6735
Trial 23 | Epoch 3/16 | Train Loss: 0.5648 | Val Loss: 0.5273 | Val Acc: 0.7551
Trial 23 | Epoch 4/16 | Train Loss: 0.5200 | Val Loss: 0.4652 | Val Acc: 0.7551
Trial 23 | Epoch 5/16 | Train Loss: 0.4614 | Val Loss: 0.4849 | Val Acc: 0.7755
Trial 23 | Epoch 6/16 | Train Loss: 0.4336 | Val Loss: 0.3699 | Val Acc: 0.8367
Trial 23 | Epoch 7/16 | Train Loss: 0.3529 | Val Loss: 0.3036 | Val Acc: 0.9082
Trial 23 | Epoch 8/16 | Train Loss: 0.2894 | Val Loss: 0.2478 | Val Acc: 0.9082
Trial 23 | Epoch 9/16 | Train Loss: 0.2410 | Val Loss: 0.3175 | Val Acc: 0.8776
Trial 23 | Epoch 10/16 | Train Loss: 0.3064 | Val Loss: 0.2570 | Val Acc: 0.9184
Trial 23 | Epoch 11/16 | Train Loss: 0.2210 | Val Loss: 0.2056 | Val Acc: 0.9388
Trial 23 | Epoch 12/16 | Train Loss: 0.1812 | Val Loss: 0.2159 | Val Acc: 0.9184
Trial 23 | Epoch 13/16 | Train Loss: 0.1510 | Val Loss: 0.2631 | Val Acc: 0.9082
Trial 23 | Epoch 14/16 | Train Loss: 0.1807 | Val Loss: 0.3579 | Val Acc: 0.8673
Trial 23 | Epoch 15/16 | Train Loss: 0.1765 | Val Loss: 0.1781 | Val Acc: 0.9286
Trial 23 | Epoch 16/16 | Train Loss: 0.1422 | Val Loss: 0.1514 | Val Acc: 0.9592
[I 2025-04-23 09:28:47,896] Trial 23 finished with value: 0.9591836734693877 and parameters: {'conv_layers': 3, 'linear_layers': 2, 'dropout_rate': 0.20137
028821881847, 'hidden_units': 512, 'batch_size': 64, 'learning_rate': 8.084076498691669e-05, 'epochs': 16}. Best is trial 23 with value: 0.9591836734693877
```

BEST RESNET CONFIGURATION

```
Trial 6 | Epoch 1/6 | Train Loss: 0.6113 | Val Loss: 0.4201 | Val Acc: 0.8163
Trial 6 | Epoch 2/6 | Train Loss: 0.4462 | Val Loss: 0.3134 | Val Acc: 0.8673
Trial 6 | Epoch 3/6 | Train Loss: 0.3158 | Val Loss: 0.2378 | Val Acc: 0.9388
Trial 6 | Epoch 4/6 | Train Loss: 0.2418 | Val Loss: 0.1845 | Val Acc: 0.9694
Trial 6 | Epoch 5/6 | Train Loss: 0.1937 | Val Loss: 0.1740 | Val Acc: 0.9490
Trial 6 | Epoch 6/6 | Train Loss: 0.1375 | Val Loss: 0.1509 | Val Acc: 0.9490
[I 2025-04-23 16:29:12,459] Trial 6 finished with value: 0.9489795918367347 and parameters: {'dropout_rate': 0.44745999926969554, 'freeze_layers': 5, 'lear
ning_rate': 1.7448710041468407e-05, 'batch_size': 16, 'epochs': 6}. Best is trial 6 with value: 0.9489795918367347.
```

RESULTS

- The fine-tuned ResNet18 model was selected and used on real-world CCTV accident videos. Each video frame was processed and classified with a threshold of 0.5.
- In successful scenarios, the model accurately detected accidents in daylight and clear conditions. The predicted labels were consistent and stable across frames.
- However, the model faced challenges in night-time or low-light conditions, where several accidents were missed due to poor visibility.
- Overall, ResNet18 proved reliable for live accident detection in normal lighting conditions. Performance can be improved with better night-time data or enhanced preprocessing techniques.



CONCLUSION

- This project demonstrated the effectiveness of deep learning models—particularly ResNet18—for real-time accident detection from CCTV footage. After extensive experimentation, ResNet18 was identified as the most suitable architecture, offering a strong balance between accuracy and efficiency.
- The system successfully classified accident scenes from live video with high reliability under normal lighting conditions. Although performance in low-light environments remains a limitation, this highlights a valuable direction for future improvement.
- Overall, the project lays a solid foundation for deploying automated accident detection systems that can support faster emergency response and enhance public safety through intelligent video surveillance.

THANK YOU