Urban Scene Cleanliness Classification using Transfer Learning and Explainable Deep learning

PREPARED BY

GOVINDA SHAW (106439)

MEDHANSH AHUJA (105982)

COURSE: LEARNING FROM IMAGES (2025-26)

UNIVERSITY: BERLINER HOCHSCHULE FÜR TECHNIK

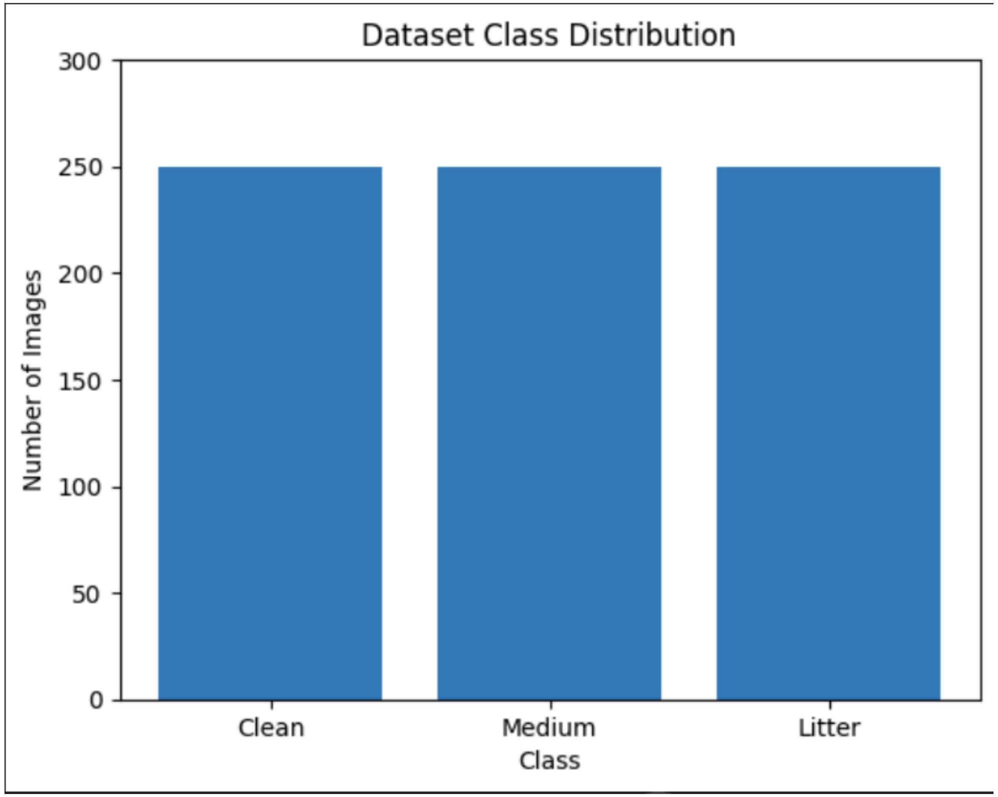
PROFESSOR : KRISTIAN HILDEBRAND

Problem statement

Urban cleanliness monitoring is important for smart city planning and public health, yet manual inspection is subjective and resource-intensive.
This project investigates whether deep convolutional neural networks can automatically classify street images into Clean, Medium, and Litter categories.
We further examine how transfer learning affects classification accuracy and decision stability in real-world urban scenes.

Dataset overview:

- Original images sourced from Kaggle dataset
- Reorganized into three classes (Clean, Medium, Litter)
- Total Images: 750
- Images per class: 250 each
- Train/Val/Test split: 70% / 15% / 15%



Sample images from each class

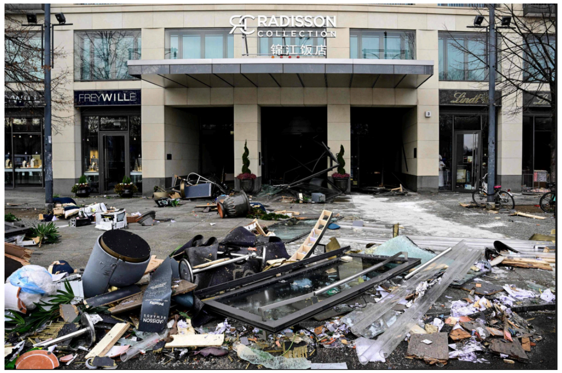
Clean



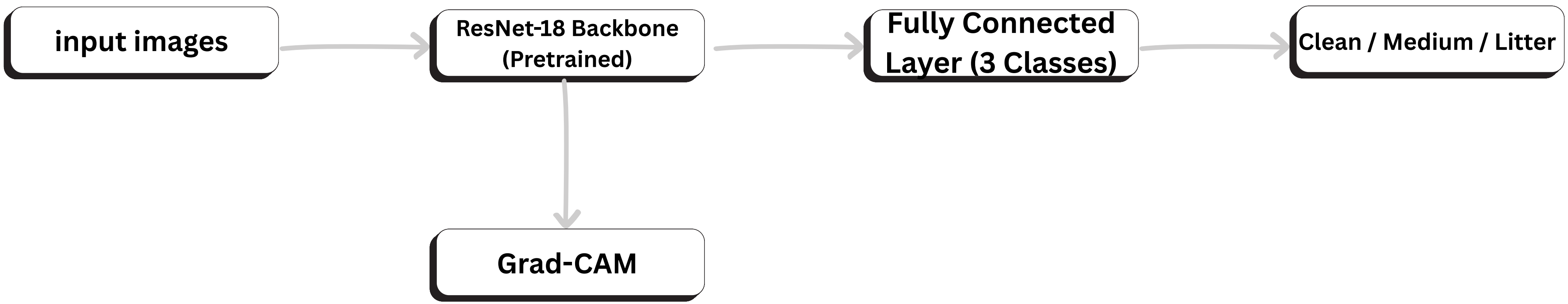
Medium




Litter:




Method overview



Model Architecture

- ResNet-18 Convolutional Neural Network
- Modified final layer for 3-class classification (Clean, Medium, Litter)

Training Strategies

- **Transfer Learning**: Pretrained on ImageNet, fine-tuned on urban dataset
- **From Scratch**: Random weight initialization, trained entirely on dataset

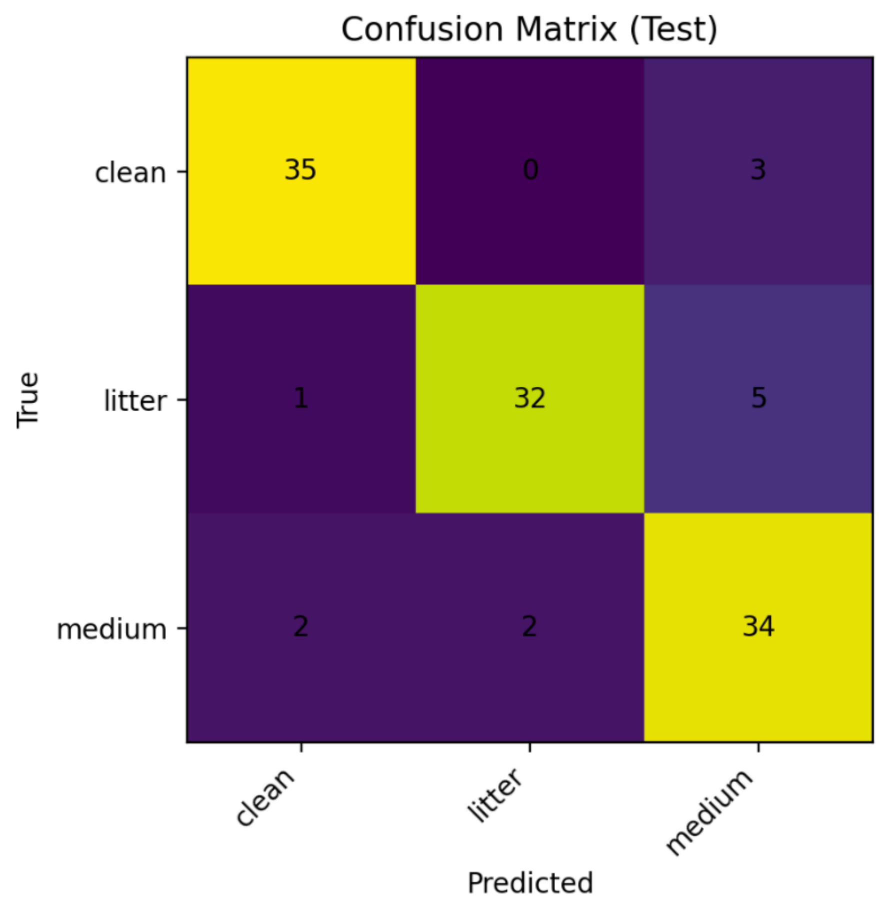
Optimization

- Loss Function: Cross-Entropy Loss
- Optimizer: Adam
- Learning rate scheduling applied

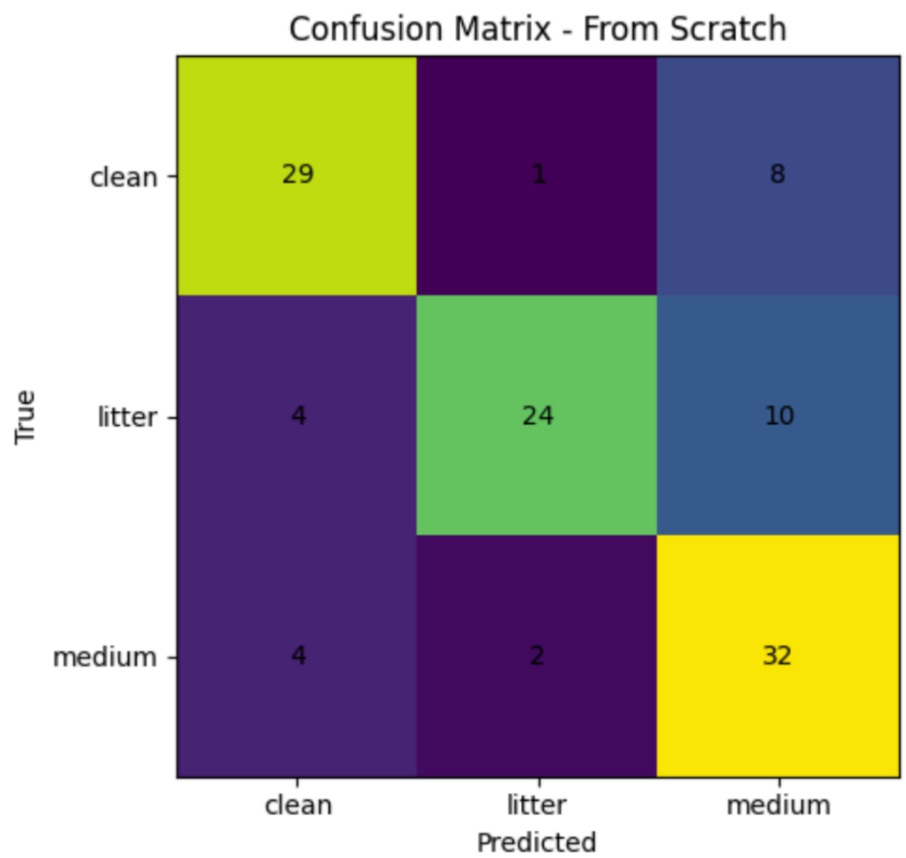
Model Interpretability

- Grad-CAM used to visualize spatial attention
- Enables analysis of decision behavior and boundary ambiguity

RESULT
confusion matrix



Pretrained:
Stronger diagonal consistency; reduced confusion and improved class discrimination.



From Scratch:
Higher confusion across classes, weaker boundary separation.

Metric Comparision

Pretrained ResNet-18

Class-wise Performance

Class	Precision	Recall	F1-Score	Support
Clean	0.921	0.921	**0.921**	38
Litter	0.941	0.842	**0.889**	38
Medium	0.81	0.895	**0.85**	38

Overall Performance

Metric	Value
Accuracy	**0.886**
Macro Precision	0.891
Macro Recall	0.886
Macro F1-score	**0.887**

From-Scratch ResNet-18

Class-wise Performance

Class	Precisio	Recall	F1-	Support
Clean	0.78	0.76	**0.77**	38
Litter	0.89	0.63	**0.74**	38
Medium	0.64	0.84	**0.73**	38

Overall Performance

Metric	Value
Accuracy	**0.75**
Macro Precision	0.77
Macro Recall	0.75
Macro F1-score	**0.75**

Model	Accuracy	Macro F1
ResNet-18 (Pretrained)	**0.886**	**0.887**
ResNet-18 (From Scratch)	0.75	0.75

Interpretation(Grad-CAM)

It answer: What is the model looking at when making decisions?

Litter-predicted as Litter

Litter-predicted as Medium

Medium-predicted as Litter

Clean-predicted as Medium

Related work

Residual Networks (He et al., 2016) significantly improved image classification through deep residual learning. Transfer learning from ImageNet has become standard practice for small datasets. However, few urban cleanliness studies evaluate interpretability alongside performance. We address this gap using Grad-CAM analysis.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. CVPR.

Selvaraju et al. (2017). Grad-CAM: Visual Explanations from Deep Networks. ICCV.