Deep Learning for Building Exterior Cladding Classification Using Pre-trained CNNs

Meltem Sahin Ozkoc (Carnegie Mellon University)

Code is available: https://github.com/meltemsahinozkoc/deep-learning-facade-classification

A. Introduction

Urban building stock classification is often labor-intensive and imprecise. The objective of this project is to explore deep learning models to classify building exterior cladding materials from street-view façade images using transfer learning. These models are based on pre-trained **ResNet50** and **InceptionV3 (GoogleNet)** architectures, applied to the **London** and **Scotland** sets of a façade image dataset. The focus is on evaluating model performance before and after augmentation, and generalization across geographic contexts.

This final report covers the data exploration, model training, and evaluation processes before and after image augmentation and tests the developed models on the Scotland dataset for a generalizability test.

B. Dataset Selection & Exploratory Data Analysis

To focus on model building rather than data collection, I selected a high-quality labeled dataset described in:
Wang, S., Park, S., Park, S., & Kim, J. (2024). Building façade datasets for analyzing building characteristics using deep learning. Data in Brief, 57, 110885.

This dataset includes exterior façade images labeled by cladding material and grouped by location (London, Scotland). The details of both datasets are given in the research article mentioned above:

The London exterior façade material dataset contains (Table 1):

- **928 training images**
- **5568 augmented images**
- **311 validation images**
- **311 test images**
- **6 classes**: Brick, Concrete, Curtain-Wall, Mixed, Others, Stone

Table 1- Detailed distribution of datasets in London (Extracted from the reference article)

Detailed distribution of datasets in London.

Purpose	Number of images	Class						
		Brick	Concrete	Glass	Stone	Mixed	Others	Total
Training	928	265	137	43	24	357	102	928
Augmentation	5568	1590	822	258	144	2142	612	5568
Validation	311	91	45	14	8	119	34	311
Test	311	86	47	15	9	120	34	311
Total	7118	2032	1051	330	185	2738	782	7118

Table 2- Definition of project target exterior façade materials classification

Definition, and visual cues of target materials.

Materials	Definition	Visual cues
Brick	Material primarily made of fired clay	Prominent masonry joints and uniform texture
Concrete	Composite material consisting of fine and coarse aggregates, cement, and water	Smooth appearance, often unpatterned, with grey to muted tones and a matte finish
Glass	Non-structural, glass-based exterior used in building facades	Glass panels without visible structural floor separations, offering reflective or transparent qualities
Stone	Natural material sourced from quarries	Natural and irregular textures, size and color variations, typically in earth tones, with finishes ranging from polished to rustic
Mixed materials	Combination of different materials used together	Two or more materials with distinguishing features, such as brick and stone, evident
Others	Any building materials not included in the above categories	Uncommon textures and appearances, including materials like wood, metal, synthetic sidings

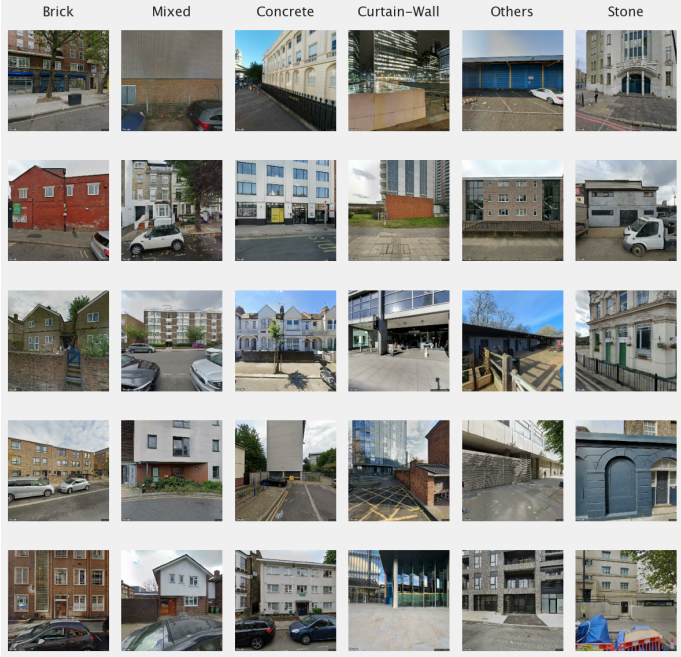


Figure 1 – Façade images from London dataset

C. Pre-trained Model Selection & Modeling Process

I employed transfer learning using:

- **ResNet50** (ImageNet weights): Chosen for its depth and residual blocks, often effective on mid-size classification tasks.
- **InceptionV3**: Known for multi-scale convolutions and strong performance in fine-grained visual tasks.

Both base models were frozen and used as fixed feature extractors, with custom classification heads added (GlobalAveragePooling → Dense(128) → Dropout(0.3) → Softmax).

I resized the images to 224×224 for ResNet50 and 299×299 for InceptionV3. Then, encoded the labels as categorical and used TensorFlow dataset pipelines for batching, shuffling, and prefetching as demonstrated in the reference course book and slides.

D. Modeling Results

Models Used:

1. **ResNet50** (frozen base + GAP + Dense(128) + Dropout(0.3) + Softmax)
2. **InceptionV3** (frozen base + GAP + Dense(128) + Dropout(0.5) + Softmax)

Training Details:

- Optimizer: Adam
- Loss: Categorical Crossentropy
- Regularizers (+Callback): EarlyStopping + ModelCheckpoint
- Training: First on non-augmented, then on augmented data

Before Augmentation

ResNet50 Performance (Unaugmented)

- `Test Accuracy: 65.92%`
- `Strength: Reasonable performance on` ***`Brick`*** `and` ***`Mixed`***
- `Weakness: Low recall on` ***`Curtain-Wall`*** `(13%) and` ***`Stone`*** `(11%)`

The ResNet50 model performed moderately well, achieving a test accuracy of 65.92%. The Brick and Mixed classes were predicted reasonably, but recall scores for underrepresented classes like Curtain-Wall and Stone were low, indicating the model struggling to generalize beyond the dominant classes. With no data augmentation and the base model frozen, the network likely lacked enough variability and depth to capture finer façade features.

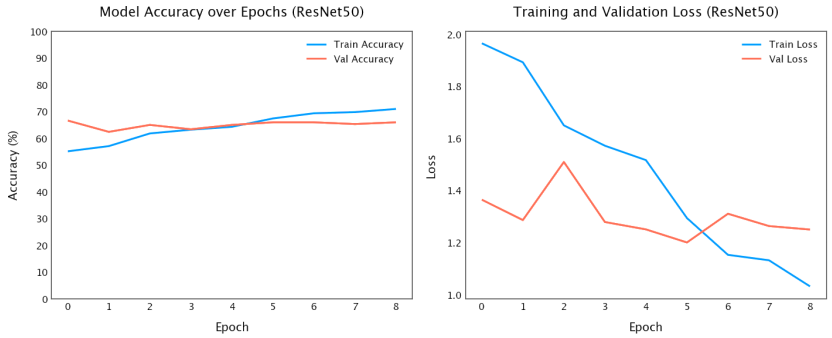


Figure 2 – Accuracy and loss plots for ResNet50 (Before augmentation)

InceptionV3 Performance (Unaugmented)

- `Test Accuracy: 68.15%`
- `Higher train accuracy but lower generalization`
- `Overfitting signs: Training loss decreases, while validation loss increases.`

In contrast, InceptionV3 reached higher training accuracy and a better test accuracy of 68.15%, but showed lower generalization. While performance on the major classes like Brick and Mixed was decent, minority classes still suffered from low recall.

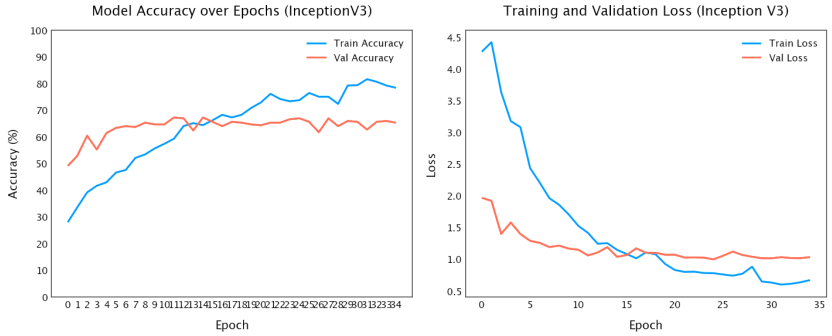


Figure 3 – Accuracy and loss plots (top) and confusion matrix (bottom) for Inception V3 (Before augmentation)

After Augmentation

The augmentation step included *brightness, contrast, perspective, rotation, scale, and shear* (Figure 4). The dataset already contained a separate folder for the augmented images. Therefore, I didn't write the script to augment the dataset I obtained.

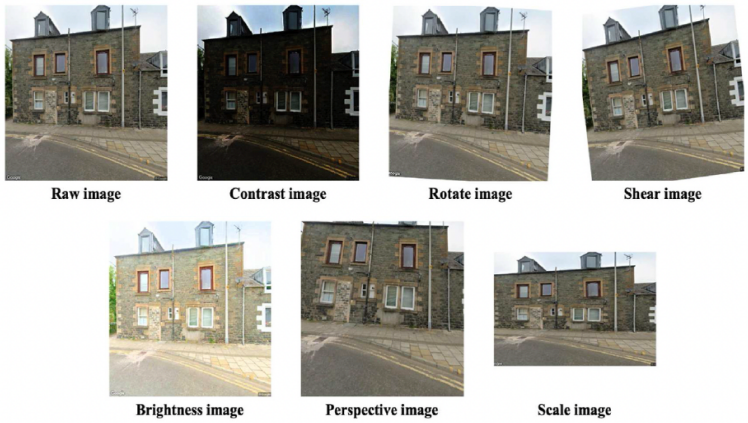


Figure 4 - Data augmentation tactics

ResNet50 Performance (Augmented)

- `Test Accuracy: 68.15%`
- `Improved balance across classes`
- `Better recall on previously weak classes`

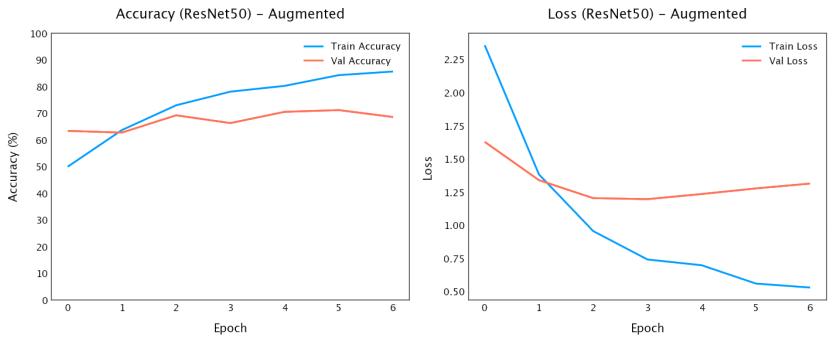


Figure 5 – Accuracy and loss plots (top) and confusion matrix (bottom) for ResNet50 (After augmentation)

InceptionV3 Performance (Augmented)

- `Test Accuracy: 70.38%`
- `Much more stable training, mitigated overfitting, than the previous unaugmented data.`
- `Most robust performance overall`

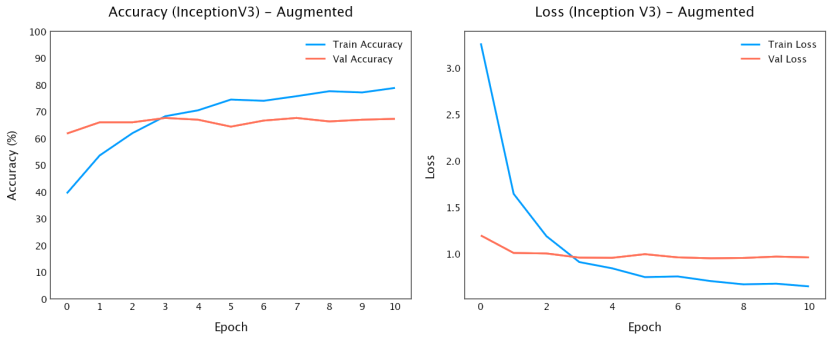


Figure 6 – Accuracy and loss plots (top) and confusion matrix (bottom) for Inception V3 (After augmentation)

E. Discussion & Insights

The main insights from the modeling process are:

- InceptionV3 outperformed ResNet50 in both training accuracy and generalization.
- Augmentation helped significantly for both models by improving recall and mitigating overfitting.
- Class imbalance and geographic generalizability remain open challenges.
- Transfer learning is a powerful tool, but its effectiveness depends on dataset variability and task complexity.

Model	*Dataset*	*Test Accuracy*	*Weighted Avg Precision*	*Weighted Avg Recall*	*Weighted Avg F1-Score*	*Macro Avg F1-Score*
ResNet50	Unaugmented	**65.92%**	66.00%	66.00%	65.00%	55.00%
InceptionV3	Unaugmented	**68.15%**	69.00%	68.00%	67.00%	54.00%
ResNet50	Augmented	**68.15%**	71.41%	68.15%	68.51%	59.57%
InceptionV3	Augmented	**70.38%**	25.00%*	26.00%*	25.00%*	16.00%*

**Interesting. I double-checked many times and am surprised to see these figures only differ for Inception.*

Model	*Dataset (Test)*	*Test Accuracy*	*Weighted Avg Precision*	*Weighted Avg Recall*	*Weighted Avg F1-Score*	*Macro Avg F1-Score*
ResNet50	Scotland	**32.21%**	54.45%	32.21%	29.61%	25.49%
InceptionV3	Scotland	**37.98%**	57.14%	37.98%	36.60%	28.85%

Testing in Scotland

The Scotland exterior façade material dataset contains (Table 1):

- **608 training images**
- **3648 augmented images**
- **204 validation images**
- **205 test images**
- **6 classes**: Brick, Concrete, Curtain-Wall, Mixed, Others, Stone

Table 2- Detailed distribution of datasets in Scotland (Extracted from the reference article)

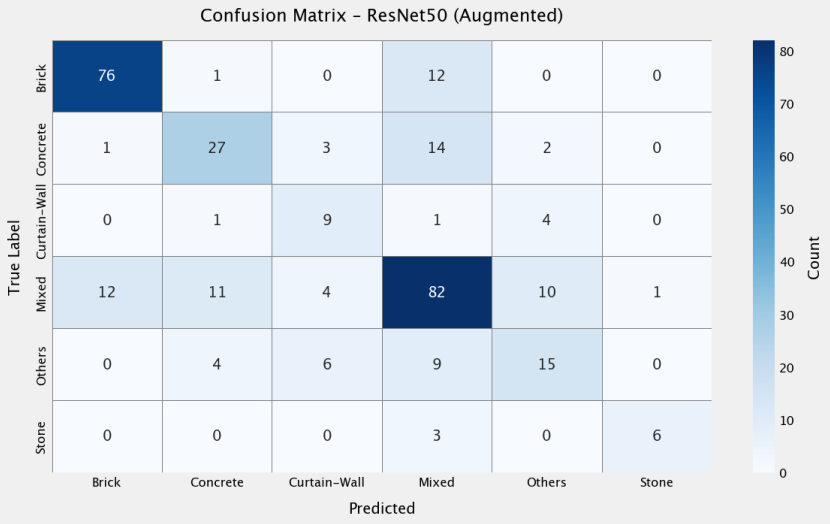
Detailed distribution of datasets in Scotland.

Purpose	Number of images	Class						
		Brick	Concrete	Glass	Stone	Mixed	Others	Total
Training	608	46	80	9	174	241	58	608
Augmentation	3648	276	480	54	1044	1446	348	3648
Validation	204	15	26	3	61	80	19	204
Test	205	16	28	4	56	81	20	205
Total	4665	353	614	70	1335	1848	445	4665

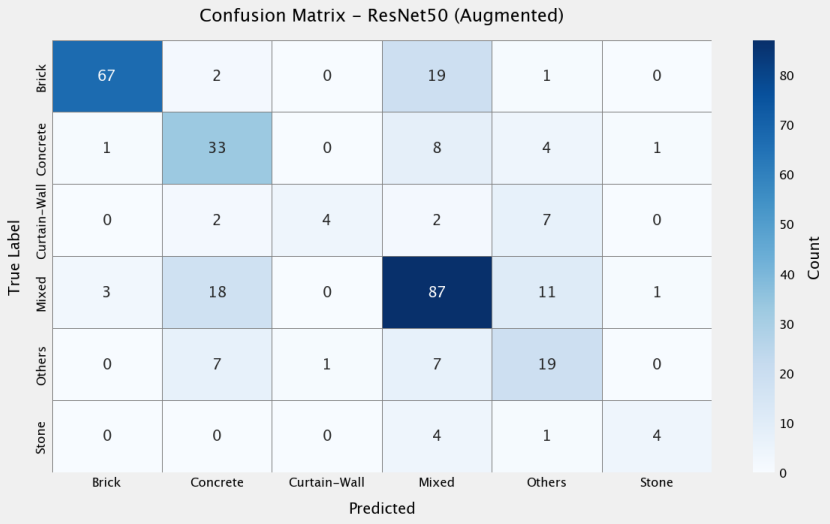
Figure 7 - Samples from each exterior façade class for Scotland

In general, the Scotland results reveal limited generalization: ResNet50 achieved 32.21% and InceptionV3 37.98% test accuracy, both with low F1-scores. This performance drop highlights a domain shift between London and Scotland images, suggesting the need for more diverse training data or a better fine-tuned model to improve cross-location robustness and generalizability.

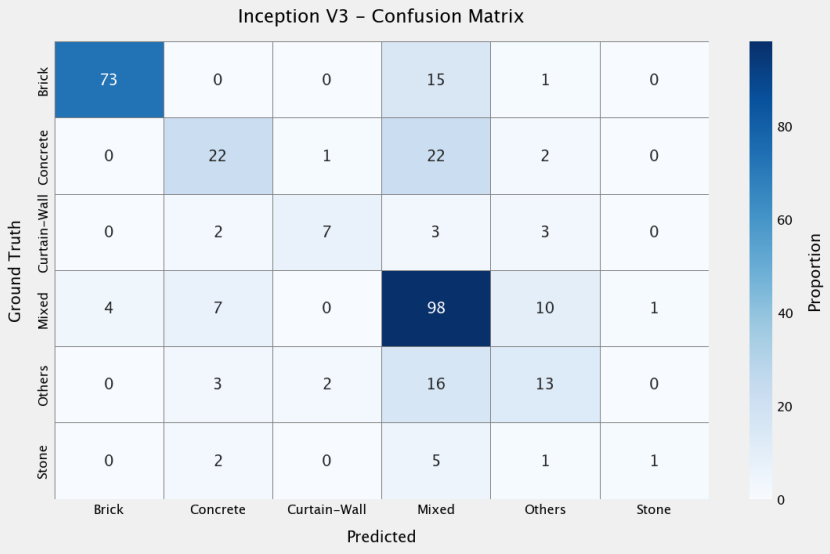
F. Appendix



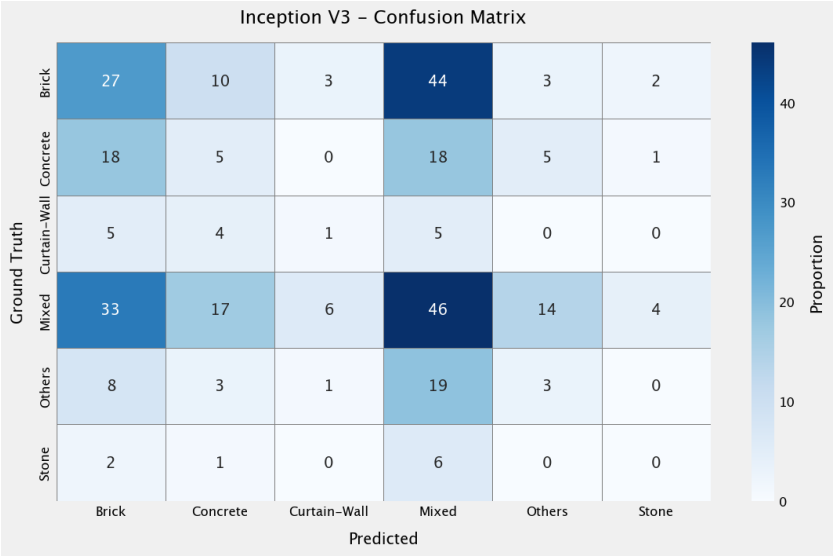
Confusion matrix for ResNet50 (Before augmentation)



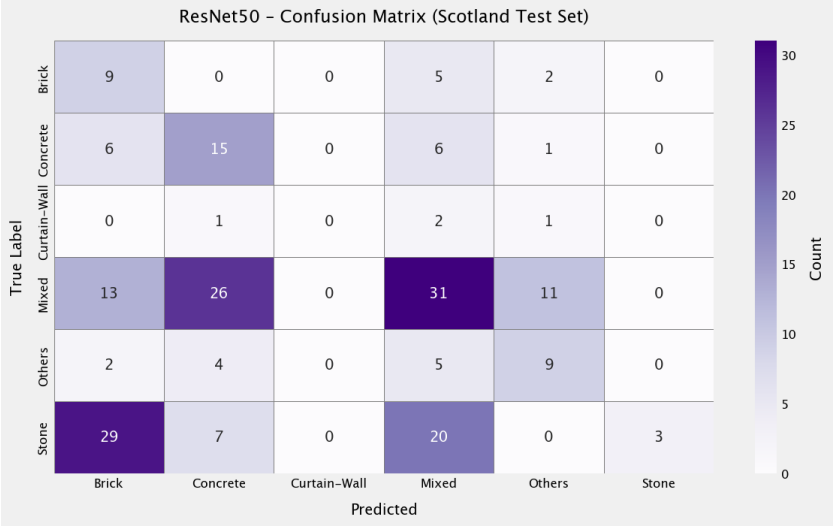
Confusion matrix for ResNet50 (After augmentation)



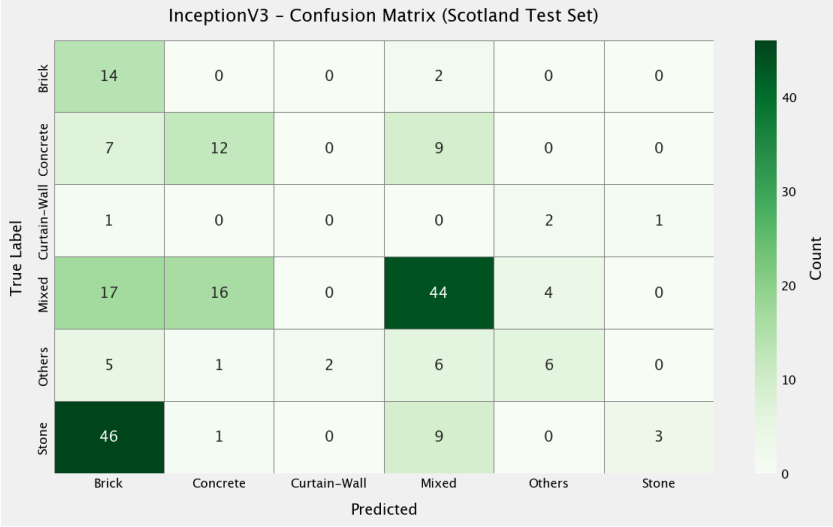
Confusion matrix for Inception V3 (Before augmentation)



Confusion matrix for Inception V3 (After augmentation)



Confusion matrix for ResNet50 (Scotland)



Confusion matrix for InceptionV3 (Scotland)