Lost in Context: The Influence of Context on Feature Attribution Methods for Object Recognition

भारतीय सांकेतिक विज्ञान संस्थ हैदराबाद
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Sayanta Adhikari*[1], Rishav Kumar*[1], Konda Reddy Mopuri[1], P. Rajalakshmi[1]

[1]Indian Institute of Technology Hyderabad

ICVGIP 2024

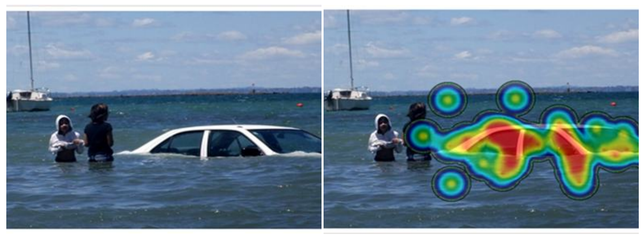
Introduction

- Contextual information significantly influences deep neural networks (DNNs) in object recognition tasks.
- Current feature attribution methods focus on objects but often fail to address the role of context.

Object : Sofa

Top 3 Predictions	
1.	Lakeside, Lakeshore
2.	Promontary, Head, Foreland
3.	Seashore, Coast

Object : Car



Top 3 Predictions	
1.	Grey whale, Devilfish
2.	Killer whale, Orca
3.	Catamaran

Figure 1. The predictions emphasize contextual information, while feature attribution for the top predicted class highlights the object in the image. Predictions were generated using a pre-trained **ResNet50** model, and feature attribution was performed with **GradCAM**.

- Investigate the influence of context manipulation on feature attribution methods and model performance.

Experimental Setup

Datasets

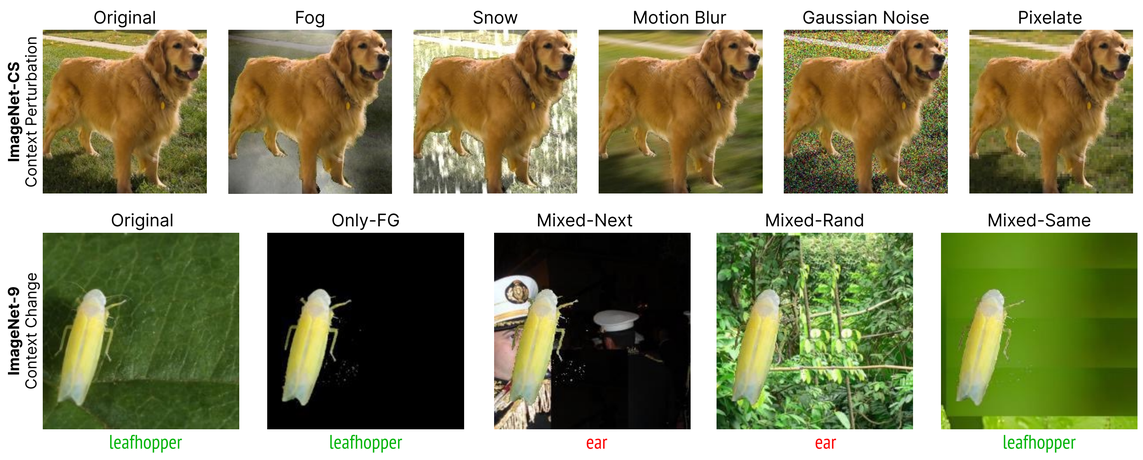


Figure 2. The top row provides images related to different varieties of our synthetic dataset named ImageNet-CS. The bottom row provides images showing variations of the ImageNet-9 dataset that we have considered for our experiments. We labelled ImageNet-9 images with its pre-trained ResNet50 classification—green, if corresponding with the original label; red, if not.

Models

- ResNet50, ResNet101, ResNet50-IN9L (trained on ImageNet-9), EfficientNet, Vision Transformer (ViT).

Feature Attribution Methods

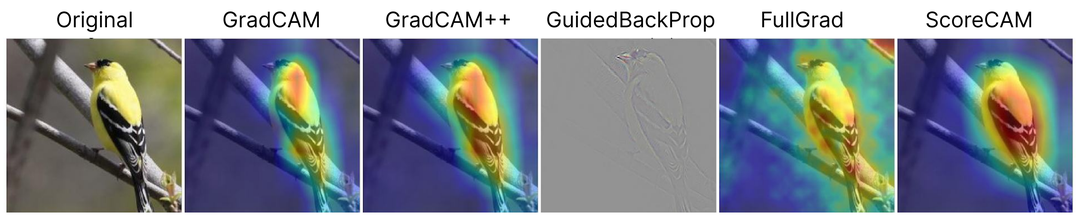


Figure 3. Comparison of feature attribution methods derived from ResNet50 classifier for top-1 prediction.

Metric

To compute attribution on object and context, we used volume attribution. We need **segmentation mask** (M_I) of the object and **feature attribution map** (A) of the image.

This simple metric gives out how much the context is important from the "eye" of a feature attribution method.

To quantify the importance of context (V_C) and object (V_O) in feature attribution maps, we define the following metric:

$$V_O = \frac{\sum(A \odot M_I)}{\sum A}, \quad V_C = \frac{\sum(A \odot (1 - M_I))}{\sum A}$$

Interpretation: This metric provides a quantitative measure to analyze how much importance the model assigns to the object and its surrounding context in an image.

Observations

a) Context Change vs. Perturbation: Which Affects Models More?

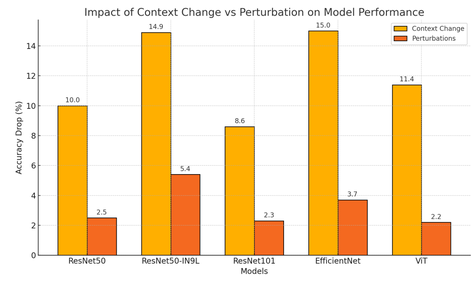
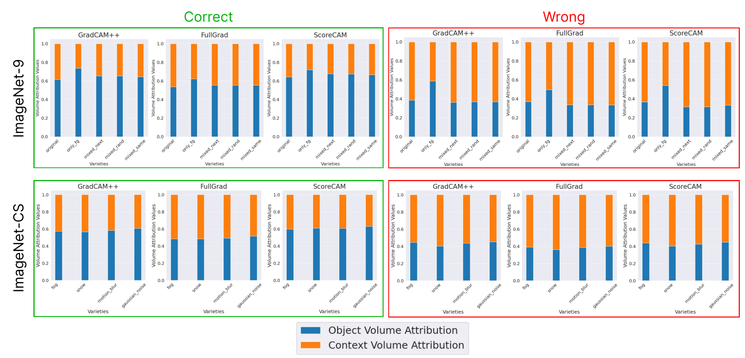


Figure 4. Impact of Context Changes vs. Perturbations on Model Performance: Context changes lead to a sharper accuracy drop across all models compared to minor perturbations

b) Does More Training Data Reduce Context Reliance?

Models trained on smaller datasets (e.g., ResNet50-IN9L) rely more on context (60%) than models trained on larger datasets (40%).

c) Why Do Models Misclassify? Context Has the Answer!



- Misclassified images show 20% higher context attribution than correctly classified ones.

Figure 5. This plot illustrates the variation in volume attribution of Context for Correctly Classified Set and Wrongly Classified Set classifications, using ResNet50 as our backbone architecture.

d) Object Size and Context: Is There a Connection?

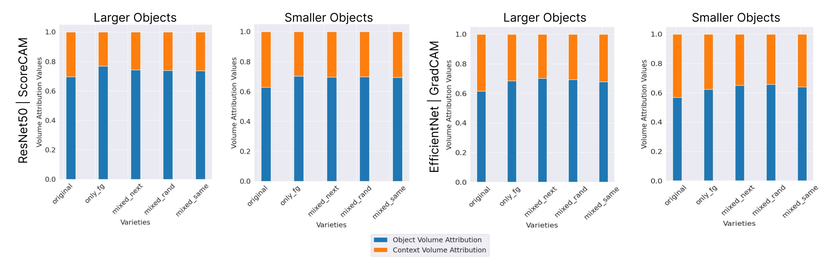


Figure 6. Average volume attribution for larger vs. smaller objects in the ImageNet-9 dataset. The plot indicates that context volume attribution is similar across object sizes, suggesting that more context pixels do not necessarily increase context influence in model predictions.

- Surprisingly, object size has little impact. Attribution remains consistent.

e) 'No Information' Context: Why Can't Models Ignore It?

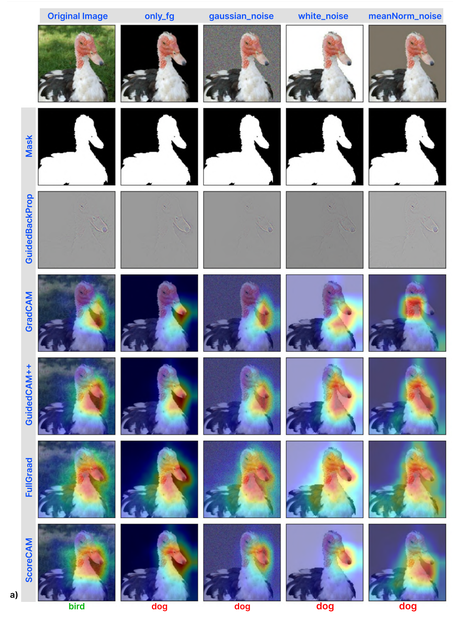


Figure 7. Qualitative Feature Attribution Comparison of 'No-Information' variants

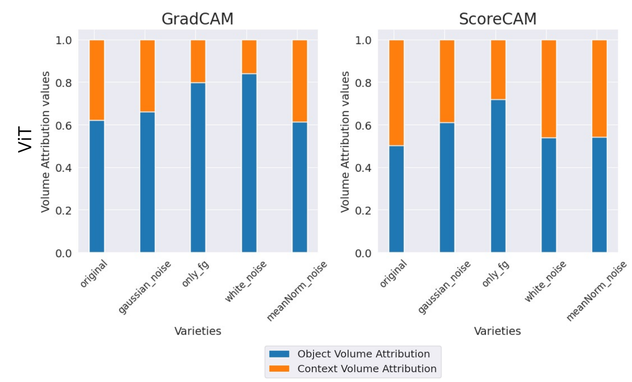


Figure 8. Feature attributions for 'no information' context varieties. Note that there is a nontrivial context volume attribution (30%) despite having 'no information' context

- Even in black or noisy backgrounds (e.g., 'only-FG', Gaussian noise), models still assign non-trivial attribution (>30%). Indicates a limitation in current feature attribution methods.

Conclusion

Key Insights:

1. **Context Changes Matter More:** Model accuracy drops significantly with context changes compared to perturbations.
2. **Training Data Size Reduces Context Reliance:** Models trained on larger datasets show less dependence on context cues.
3. **Misclassification and Context:** Higher context attribution correlates with misclassified predictions.
4. **Object Size is Not the Key:** Context attribution remains consistent regardless of object size.
5. **The 'No Information' Paradox:** Even black or noisy backgrounds attract non-trivial attributions, revealing limitations in current feature attribution methods.

Takeaway: Our findings highlight the need for models that prioritize intrinsic object features over contextual shortcuts. Additionally, feature attribution methods must be refined to accurately distinguish relevant and irrelevant context, enhancing model robustness and explainability.