

# Machine Learning - HA12

Weili Diao Student ID: 21127071

Program: Big Data Techonology Email: wdiaoaa@connect.ust.hk

## 1 Introduction

### 1.1 Project Objective

The task of this assignment is to generate explanations for the classification results of 10 test images using two different Convolutional Neural Network (CNN) models pre-trained on ImageNet. The Grad-CAM and LIME explanation methods will be employed to visualize and analyze the models' predictions.

This report begins by outlining the experimental setup. It then presents the predictive scores for each image and heatmaps generated using Grad-CAM and LIME for the predictions . Finally, a comparative discussion is conducted.

### 1.2 Preparation

#### 1.2.1 Models

The two models used in this experiment are ResNet-50 (models.resnet50(pretrained=True)) and VGG16 (models.vgg16(pretrained=True)).

#### 1.2.2 Environment

Google Colab: Python 3 CPU

Python version: 3.10.12

Operating System: Linux 6.1.85+

PyTorch version: 2.5.1+cu121

## 2 Experiment Results

### 2.1 Picture 1: golfcart

#### 2.1.1 ResNet-50

most likely class: golfcart 67.88%

Grad-CAM (Figure 1):

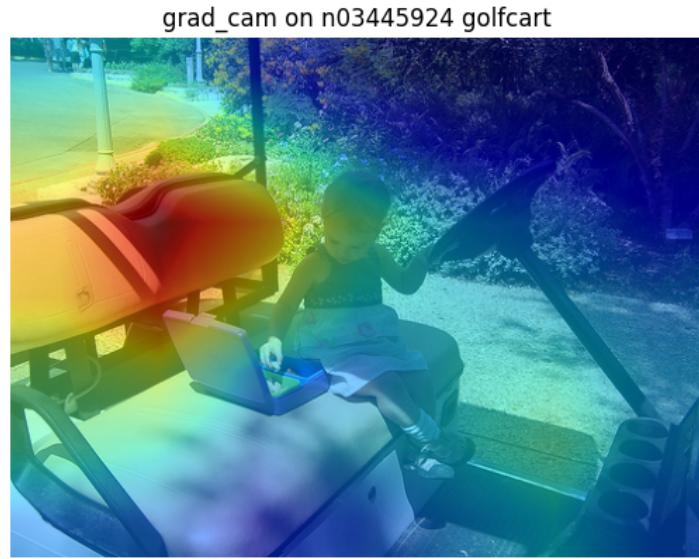


Figure 1

LIME (Figure 2), note that num\_features=5 for the first one and num\_features=10 for another, and the same below:



Figure 2

### 2.1.2 VGG16

most likely class: golfcart 98.50%  
 Grad-CAM (Figure 3):

grad\_cam on n03445924 golfcart

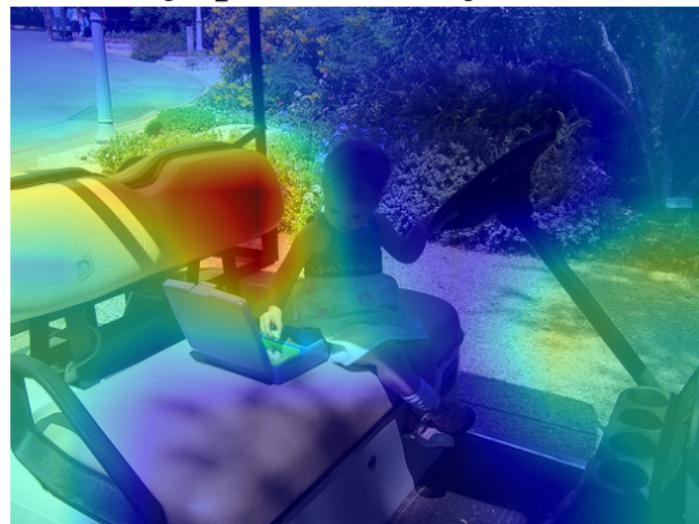


Figure 3

LIME (Figure 4):

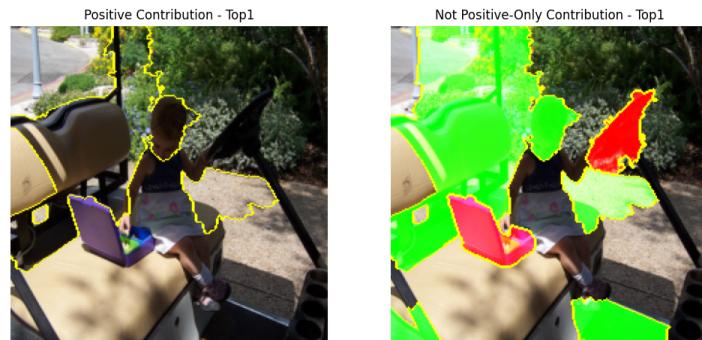


Figure 4

## 2.2 Picture 2: water bottle

### 2.2.1 ResNet-50

most likely class: syringe 53.87%

Grad-CAM (Figure 5):

grad\_cam on n04376876 syringe



Figure 5

LIME (Figure 6):

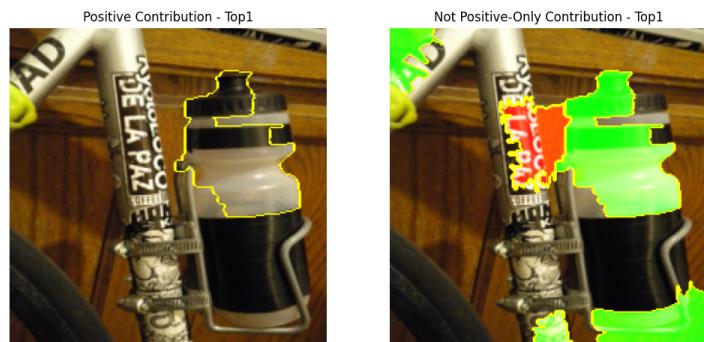


Figure 6

### 2.2.2 VGG16

most likely class: vacuum 53.34%

Grad-CAM (Figure 7):

grad\_cam on n04517823 vacuum



Figure 7

LIME (Figure 8):

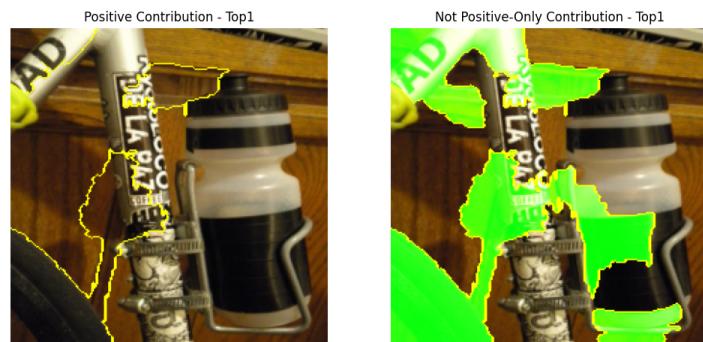


Figure 8

## 2.3 Picture 3: library

### 2.3.1 ResNet-50

most likely class: bookshop 50.53%

Grad-CAM (Figure 9):

grad\_cam on n02871525 bookshop

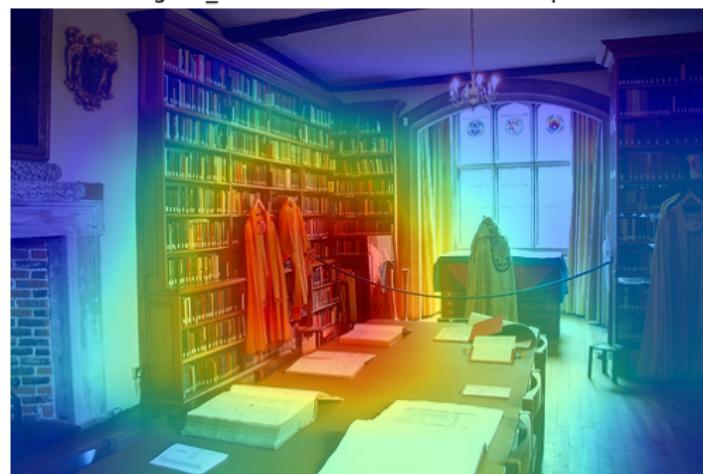


Figure 9

LIME (Figure 10):



Figure 10

### 2.3.2 VGG16

most likely class: library 94.92%  
Grad-CAM (Figure 11):



Figure 11

LIME (Figure 12):

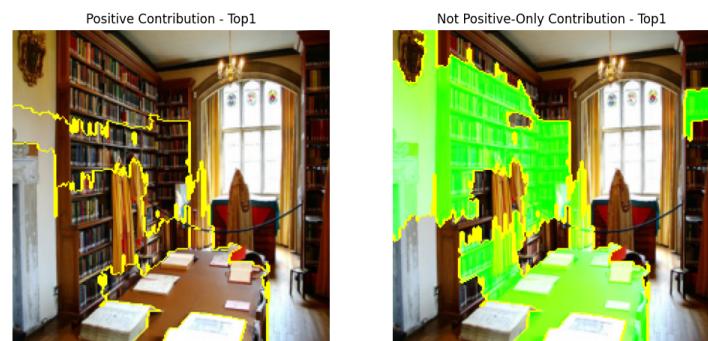


Figure 12

## 2.4 Picture 4: car mirror

### 2.4.1 ResNet-50

most likely class: car mirror 99.51%  
Grad-CAM (Figure 13):

grad\_cam on n02965783 car\_mirror



Figure 13

LIME (Figure 14):

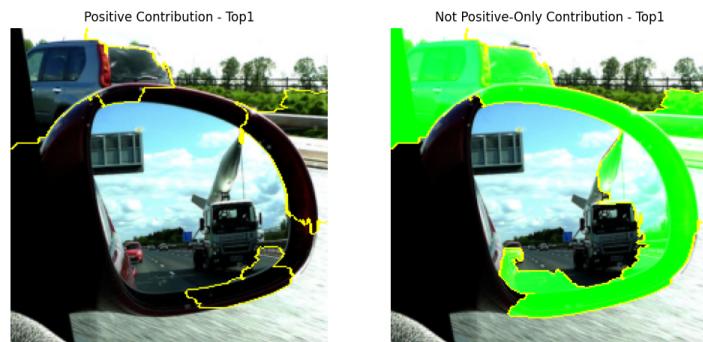


Figure 14

#### 2.4.2 VGG16

most likely class: car mirror 99.67%

Grad-CAM (Figure 15):

grad\_cam on n02965783 car\_mirror



Figure 15

LIME (Figure 16):



Figure 16

## 2.5 Picture 5: desktop computer

### 2.5.1 ResNet-50

most likely class: laptop 23.47%

Grad-CAM (Figure 17):



Figure 17

LIME (Figure 18):

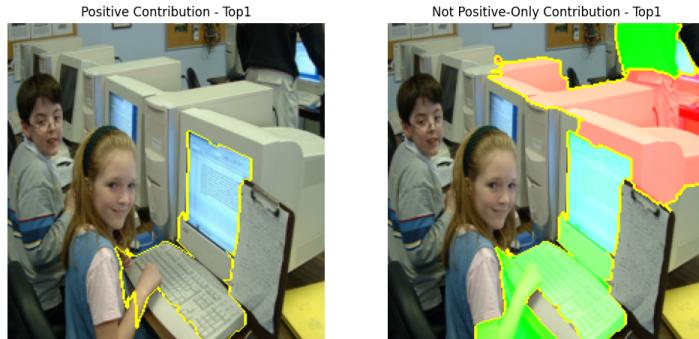


Figure 18

### 2.5.2 VGG16

most likely class: laptop 42.42%

Grad-CAM (Figure 19):



Figure 19

LIME (Figure 20):

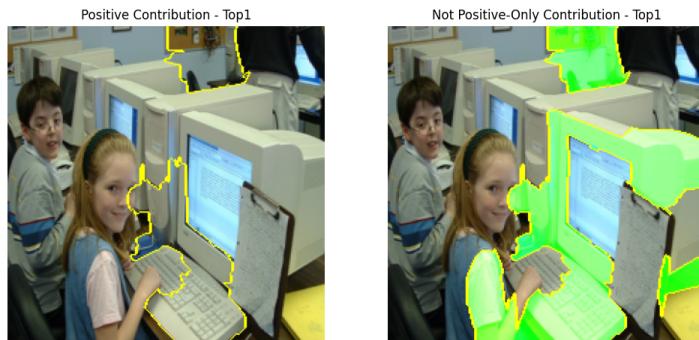


Figure 20

## 2.6 Picture 6: zebra

### 2.6.1 ResNet-50

most likely class: zebra 99.99%

Grad-CAM (Figure 21):



Figure 21

LIME (Figure 22):

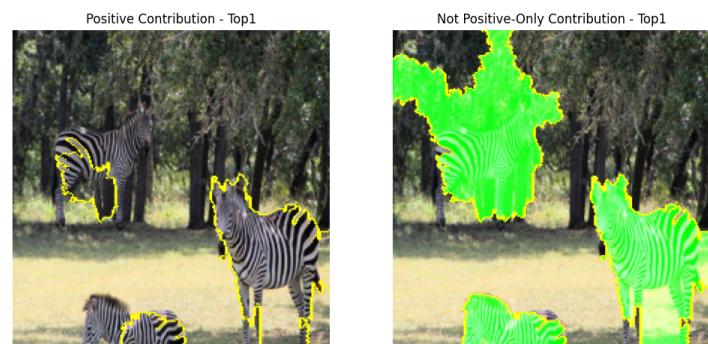


Figure 22

### 2.6.2 VGG16

most likely class: zebra 100.00%

Grad-CAM (Figure 23):



Figure 23

LIME (Figure 24):

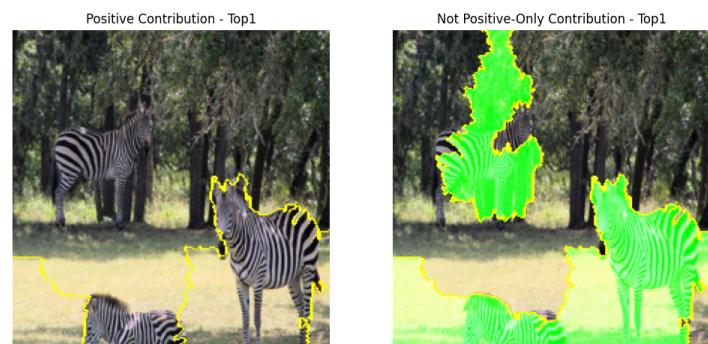


Figure 24

## 2.7 Picture 7: school bus

### 2.7.1 ResNet-50

most likely class: school bus 100.00%  
Grad-CAM (Figure 25):

grad\_cam on n04146614 school\_bus



Figure 25

LIME (Figure 26):



Figure 26

### 2.7.2 VGG16

most likely class: school bus 100.00%

Grad-CAM (Figure 27):

grad\_cam on n04146614 school\_bus



Figure 27

LIME (Figure 28):



Figure 28

## 2.8 Picture 8: pillow

### 2.8.1 ResNet-50

most likely class: pillow 100.00%

Grad-CAM (Figure 29):

grad\_cam on n03938244 pillow



Figure 29

LIME (Figure 30):

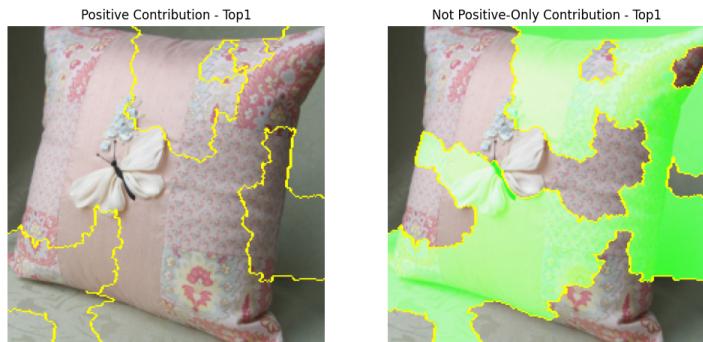


Figure 30

### 2.8.2 VGG16

most likely class: pillow 100.00%

Grad-CAM (Figure 31):

grad\_cam on n03938244 pillow

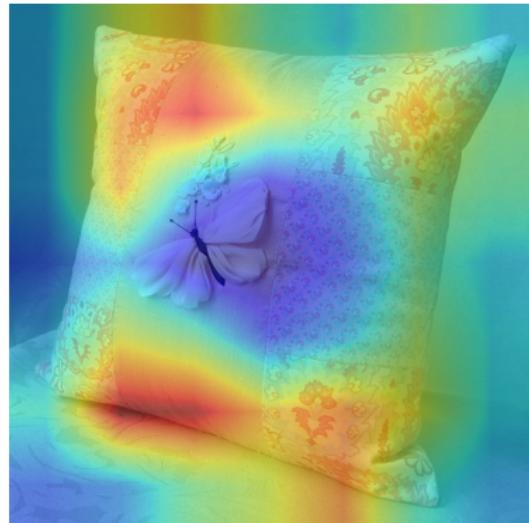


Figure 31

LIME (Figure 32):

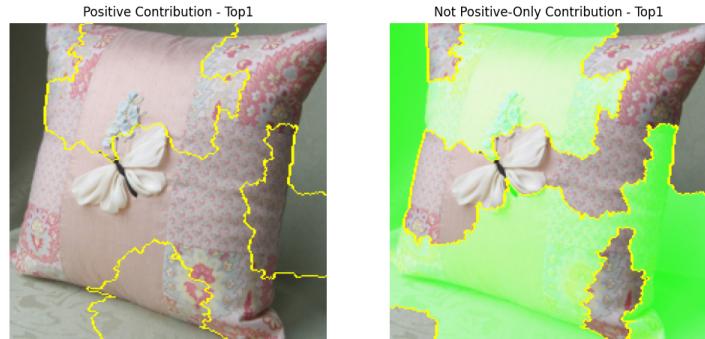


Figure 32

## 2.9 Picture 9: fireboat

### 2.9.1 ResNet-50

most likely class: fireboat 100.00%

Grad-CAM (Figure 33):

grad\_cam on n03344393 fireboat

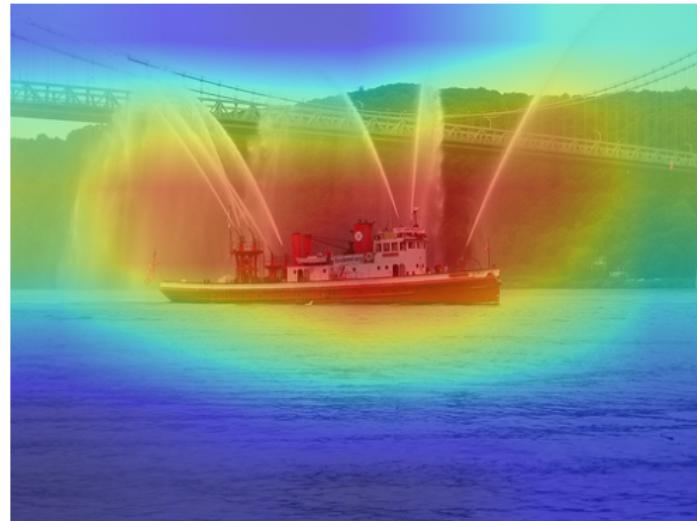


Figure 33

LIME (Figure 34):

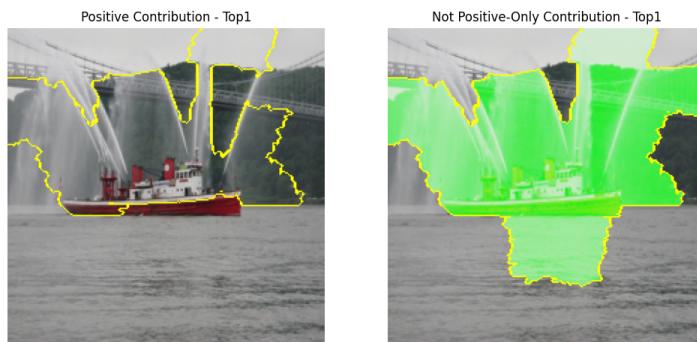


Figure 34

### 2.9.2 VGG16

most likely class: fireboat 100.00%

Grad-CAM (Figure 35):

grad\_cam on n03344393 fireboat

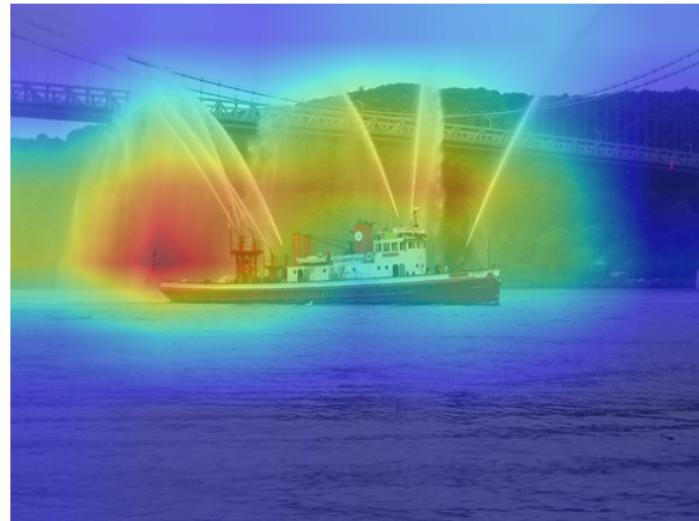


Figure 35

LIME (Figure 36):

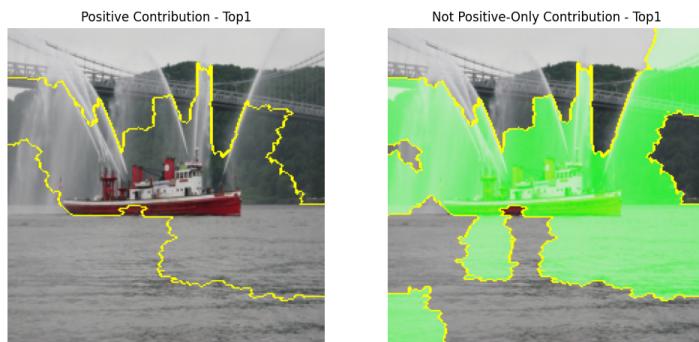


Figure 36

## 2.10 Picture 10: carousel

### 2.10.1 ResNet-50

most likely class: carousel 100.00%

Grad-CAM (Figure 37):

grad\_cam on n02966193 carousel



Figure 37

LIME (Figure 38):



Figure 38

### 2.10.2 VGG16

most likely class: carousel 100.00%

Grad-CAM (Figure 39):



Figure 39

LIME (Figure 40):



Figure 40

### 3 Discussion on the Results

#### 3.1 Accuracy

Firstly, in terms of the classification accuracy of the models. For the ResNet-50 model, misclassifications occurred for Picture 2, 3, and 5. In contrast, the VGG16 model made errors in classifying Picture 2 and 5.

#### 3.2 Heatmaps

Next, we compare the heatmaps generated by Grad-CAM and LIME. The heatmaps produced by Grad-CAM typically highlight smooth and continuous regions in the image, with areas having

greater influence on the prediction shown in warmer colors like red, while those having smaller influence shown in colder colors like purple. On the other hand, LIME method generates two types of visualizations. In the first image, the outlined regions indicate positive contributions to the prediction. In the second image, green and red patches respectively represent positive and negative contributions, illustrating the areas that support or detract from the model’s classification decision. This difference arises because the LIME method uses a perturbation-based approach, which segments the image into superpixels and evaluates how each region contributes to the model’s output.

### 3.3 Example: Picture 3

Next, let us first discuss the classification results and reasons for the Picture 3. ResNet-50 misclassified this picture, while VGG16 correctly identified it.

ResNet-50 misclassified the picture as "bookshop." The Grad-CAM heatmap reveals that the model predominantly focuses on the bookshelves and the central table and architectural features are largely ignored. In the LIME heatmap, the positive highlighted regions also correspond to the bookshelves and books, and the negative contributions highlight parts of the architectural structure, which is significant in differentiating a library from a bookshop.

VGG16 correctly classified the image as "library." Its Grad-CAM heatmap shows a more distributed focus, capturing both the books and architectural features of a library. The LIME heatmap as well supports this observation, with positive contributions distributed across bookshelves and the table.

From above we can conclude that the narrow attention of ResNet-50 likely leads to the misclassification as "bookshop.". ResNet-50's reliance on a single dominant feature (bookshelves) made it vulnerable to environment ambiguities, while VGG16's comprehensive approach allowed for more accurate classification.

### 3.4 Example: Picture 5

We can notice that both models misclassified Picture 5, and the probability predicted by ResNet-50 for "laptop" was very low, at only 23.47%. Therefore, let us further discuss the heatmaps for Picture 5.

The Grad-CAM heatmap for ResNet-50 shows a scattered focus on the screen and keyboard, lacking clear attention to defining features of a laptop. LIME shows the same, with positive contributions on the screen and keyboard while negative contributions on the monitor of computer.

VGG16's Grad-CAM heatmap is also concentrated on the keyboard, but it overlooks elements like the monitor and classroom setting. LIME indicates similar behavior, with positive contributions on laptop-like features but failing to incorporate the broader environment into its decision-making.

For this picture, both models misclassified it as "laptop" due to overemphasis on partial features like the screen and keyboard.