Manipulating images for more misclassifications:

To intentionally increase the misclassification rate of a Convolutional Neural Network (CNN) on a test dataset, I employed various techniques, such as applying different filters, adjusting normalization parameters, and introducing noise. One particular technique I used was called Adversarial Attack, which involves adding carefully crafted noise to an image in a way that the human eye cannot detect, but which causes the CNN to misclassify the image.

I followed a two-step process in my approach. Firstly, I applied various filters to the images to manipulate them in different ways. Then, in the second step, I added adversarial noise to some of the filtered images.

Interestingly, I found that the effect of adding noise or adversarial attack was not the same for all images. For some filtered images, adding noise increased the rate of misclassification, while for others, adding adversarial attack decreased the rate of misclassification.

Test Dataset:49

Training Dataset: 635

In the test dataset, ResNET50 misclassified 18 out of 49 images without any filters or manipulation.

To summarize, ResNet50 misclassified 47 out of 49 test dataset images when using the Contour filter without any noise. After adding adversarial noise, the GaussianBlur filter caused the most misclassifications, with 47 out of 50 test dataset images being misclassified. Before adding noise, the maximum number of misclassifications was caused by the Contour filter, while after adding noise, the GaussianBlur filter caused the most misclassifications.

Below, you can find detailed results of the effects of applying filters and adding adversarial noise to each filtered image on the misclassification rate of ResNet50:

1- Adverserial attack

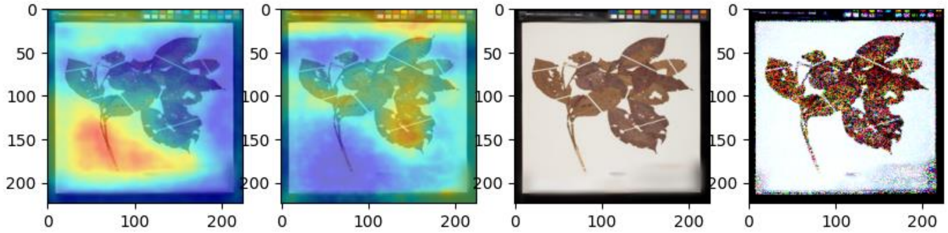
Number of misclassifications: 40

Adversarial attack is a technique used to intentionally perturb an image in a way that makes it difficult for a Convolutional Neural Network (CNN) to accurately classify it. The perturbations added to the image are typically imperceptible to the human eye, but they cause the CNN to misclassify the image with a high degree of confidence.

Adversarial attacks can take various forms, but one common approach is to use an optimization algorithm to find the minimal perturbation needed to cause a misclassification. This perturbation is then added to the original image to create the adversarial image. The goal of this technique is to identify weaknesses in the CNN's decision boundary and to test its robustness to different kinds of attacks.

The effect of adversarial attacks on misclassification can be significant. Adversarial attacks can cause a CNN to misclassify an image with high confidence, even if the image is very similar to another image that the CNN correctly classified. This can have serious consequences in applications such as autonomous driving or medical imaging, where misclassification can lead to life-threatening situations.

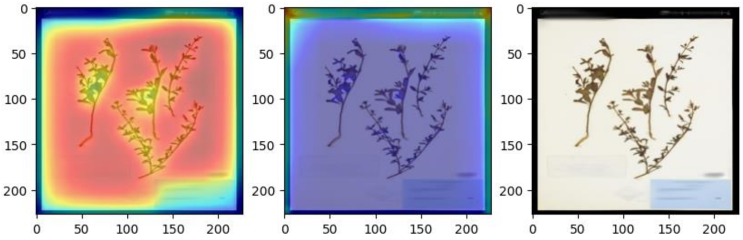
However, it is important to note that adversarial attacks can also be used to improve the robustness of a CNN by identifying weaknesses and areas for improvement. By analyzing the patterns and types of perturbations that cause misclassification, researchers can develop new techniques to improve the CNN's accuracy and resilience to attacks.



2- Change normalize parameters:

Number of misclassifications: 30

Normalizing is a common technique used in machine learning to scale and normalize input data, including in CNNs. By adjusting normalization parameters, the way a CNN detects and learns features can be altered, potentially affecting the misclassification rate. Changing the normalization parameters can improve or reduce the accuracy of a CNN, and it is crucial to evaluate the effects carefully.

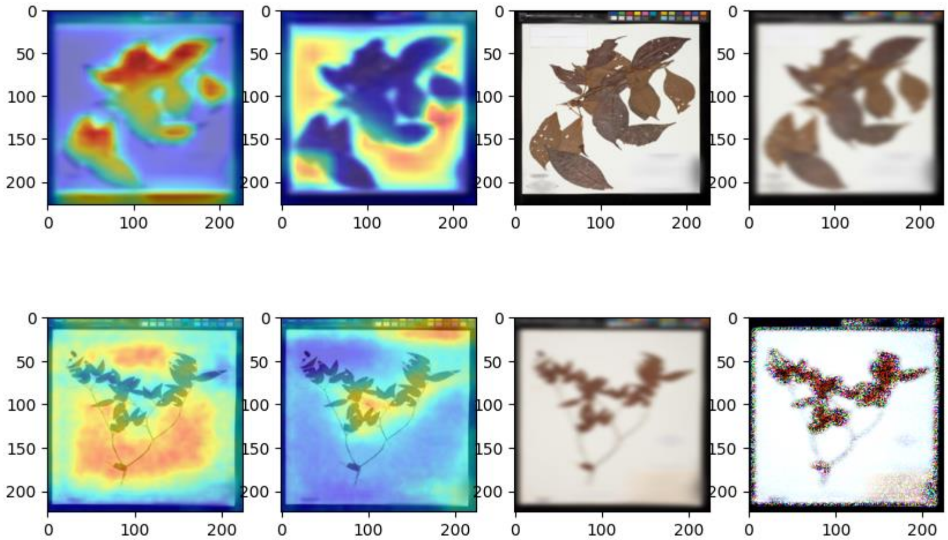


3- GaussianBlur(radius=10)

Number of misclassifications: 35

Number of misclassifications after adding noise: 47

GaussianBlur is a filter used to smooth and blur images, and it can be applied to modify the input data to a CNN. The effect of GaussianBlur on the misclassification rate of a CNN depends on the level of blurring and the specific dataset. In some cases, GaussianBlur can improve accuracy by reducing image noise, while in other cases, it can decrease accuracy by removing too much detail.

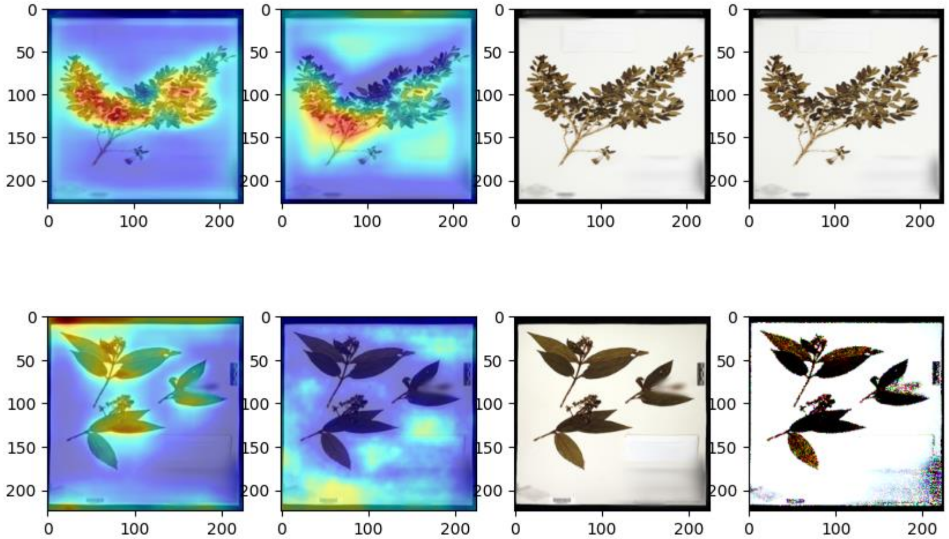


4- Sharpen: makes edges of objects more distinct

Number of misclassifications: 18

Number of misclassifications after adding noise: 40

Sharpen is a filter used to increase the contrast and edges of an image, and it can be applied to modify the input data to a CNN. The effect of Sharpen on the misclassification rate of a CNN depends on the level of sharpening and the specific dataset. In some cases, adding Sharpen can improve accuracy by enhancing edges and details, while too much Sharpen can introduce artifacts that hinder accurate classification.

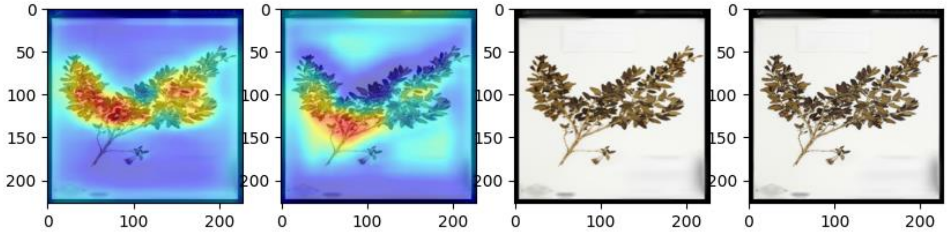


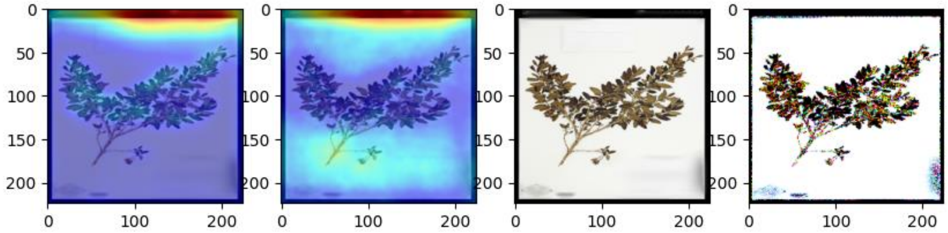
5- Edge Enhance: emphasizes the edges in an image

Number of misclassifications: 18

Number of misclassifications after adding noise: 40

Edge Enhance filters can be used to modify the input data to a CNN by enhancing the edges and details of an image. Their effect on the misclassification rate of a CNN depends on the specific dataset and level of enhancement applied. Edge Enhance filters can improve accuracy by highlighting the edges and details of the image, but too much enhancement can lead to misclassification. Adding noise to an enhanced filtered image can increase the misclassification rate of a CNN by introducing new features and patterns that the CNN may misinterpret.



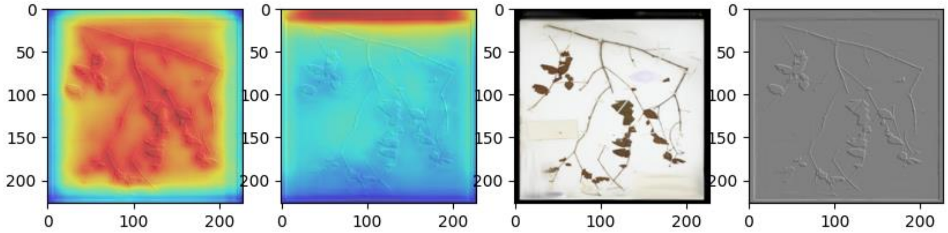


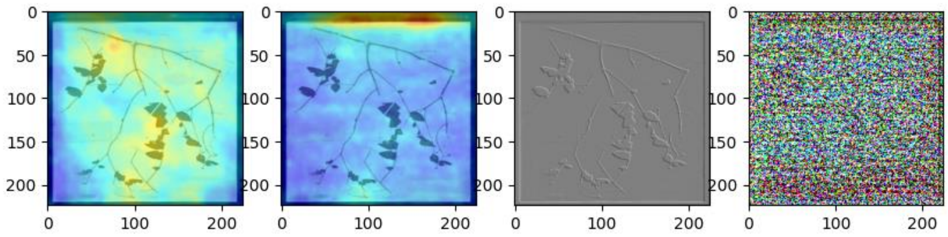
6- Emboss: adds an embossed effect to the image

Number of misclassifications: 38

Number of misclassifications after adding noise: 46

Embossed filters are image filters that enhance the edges of an image to create a 3D effect, and applying them to an image can potentially affect the misclassification rate of a CNN. Excessive use of embossed filters can lead to misclassification. Adding noise to embossed filtered images can increase the misclassification rate of a CNN.



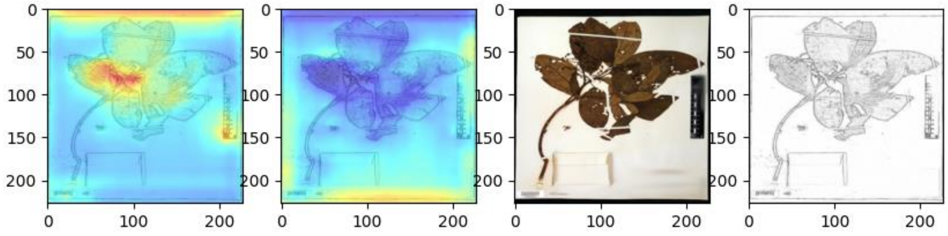


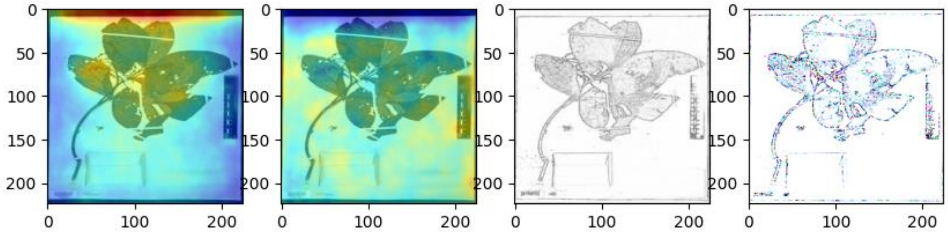
7- Contour: applies a contour filter to the image

Number of misclassifications: 47

Number of misclassifications after adding noise: 46

Applying contour filters to an image can potentially increase the misclassification rate of a CNN. Contour filters are image filters that enhance the boundaries or edges of objects in an image. This can make it easier for the CNN to distinguish between different objects or classes, but it can also introduce additional features or patterns that the CNN may misinterpret, leading to an increase in misclassification, but here contour filters can lead to more misclassification.



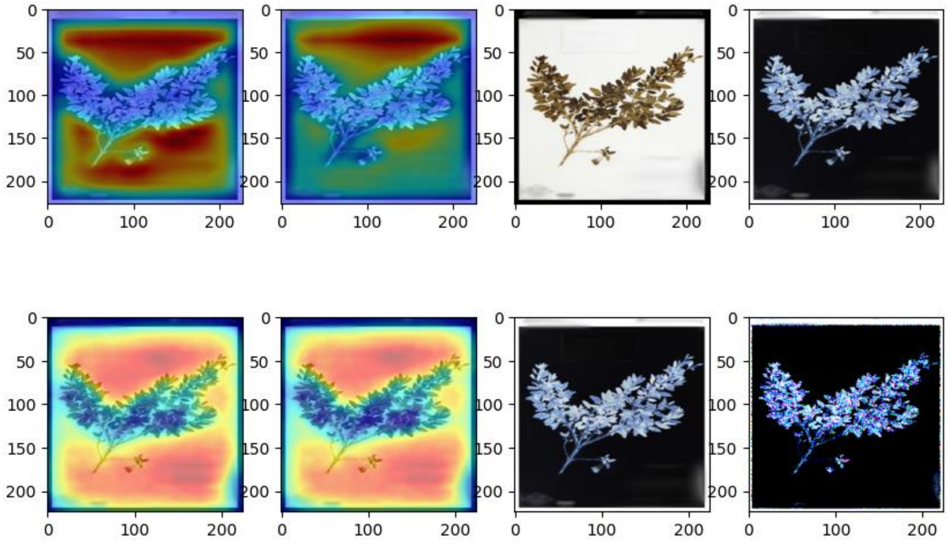


8- Invert: inverts the colors of the image

Number of misclassifications: 39

Number of misclassifications after adding noise: 38

Applying an invert filter to an image can potentially increase the misclassification rate of a CNN. An invert filter is a type of image filter that changes the colors of an image, such that the brighter parts become darker and vice versa. This can make it more difficult for the CNN to distinguish between different objects or classes, especially if the color scheme of the objects is an important feature for classification.

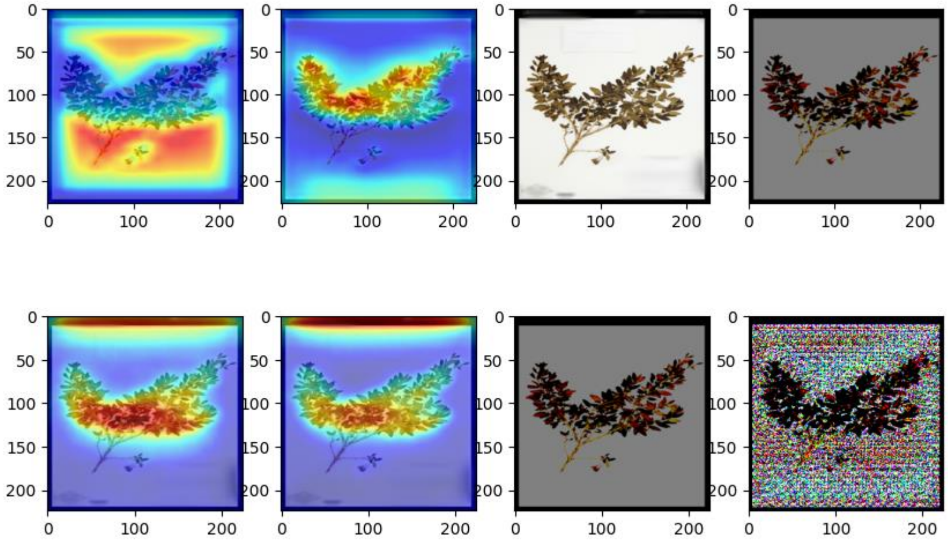


9- Posterize: reduces the number of colors in the image

Number of misclassifications: 46

Number of misclassifications after adding noise: 38

Applying a posterize filter to an image can potentially increase the misclassification rate of a CNN. A posterize filter is a type of image filter that reduces the number of colors in an image by grouping similar colors together. This can make it more difficult for the CNN to distinguish between different objects or classes, especially if the color scheme of the objects is an important feature for classification.

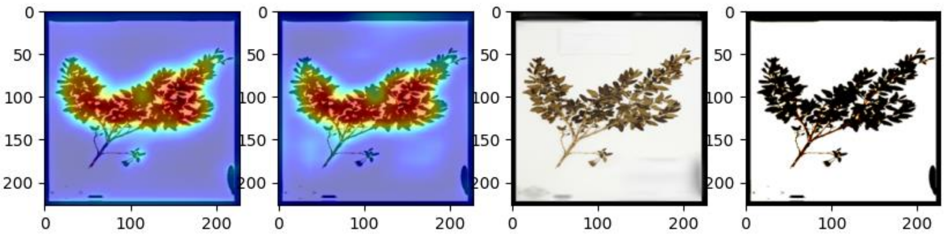


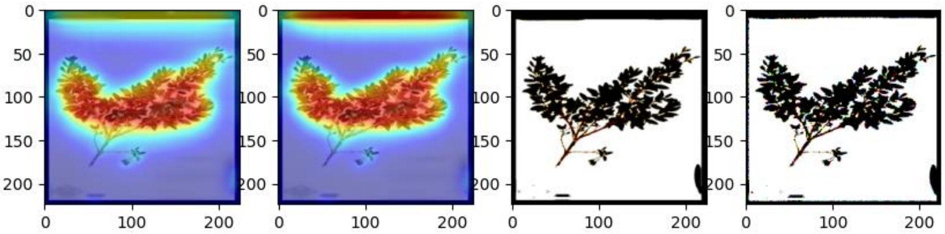
10- Contrast: adjusts the contrast of the image

Number of misclassifications: 46

Number of misclassifications after adding noise: 38

Using contrast filters to adjust the brightness and contrast of an image can potentially lead to an increased misclassification rate of a CNN. Contrast filters are image filters that adjust the brightness and contrast of an image to enhance its visual quality. While this can make the image more visually appealing to humans, it can also introduce additional noise and artifacts that can be confusing for a CNN.





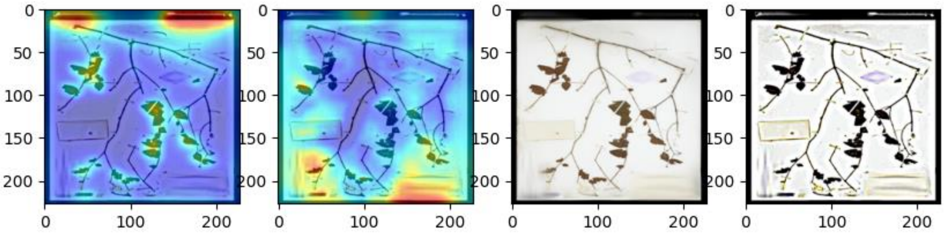
11- Unsharp Mask Filter

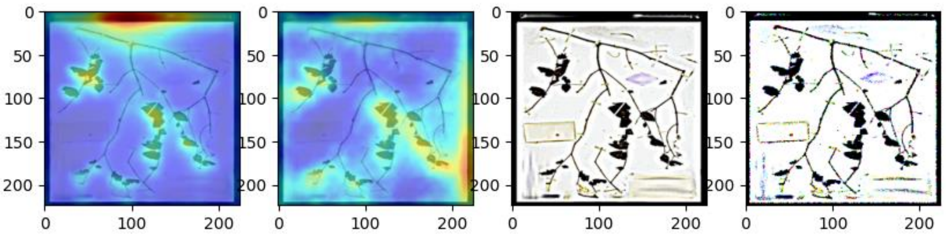
Number of misclassifications: 39

Number of misclassifications after adding noise: 38

The Unsharp Mask Filter, which is used to sharpen edges and details in an image, was found to increase the misclassification rate for the ResNet50 CNN. This filter works by creating a blurred version of the image, subtracting this from the original image to create an "edge" image, and then adding the edge image back to the original to enhance its details.

The addition of the edge image can introduce noise and artifacts that can confuse the model and cause misclassification.



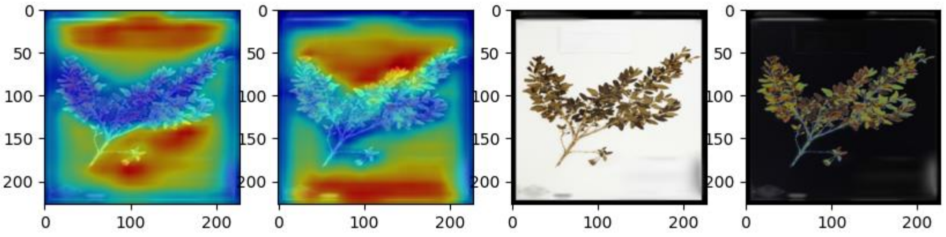


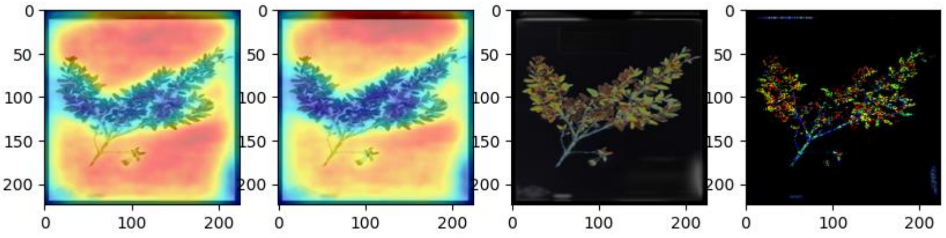
12- Solarize

Number of misclassifications: 41

Number of misclassifications after adding noise: 38

The Solarize filter, which inverts the pixel values of an image above a certain threshold, was found to increase the misclassification rate of a CNN. This filter works by thresholding the image at a certain value and then inverting the pixel values above that threshold. The application of the Solarize filter can result in the introduction of unwanted noise and artifacts in an image. This can lead to confusion for the CNN and ultimately cause a higher rate of misclassification.





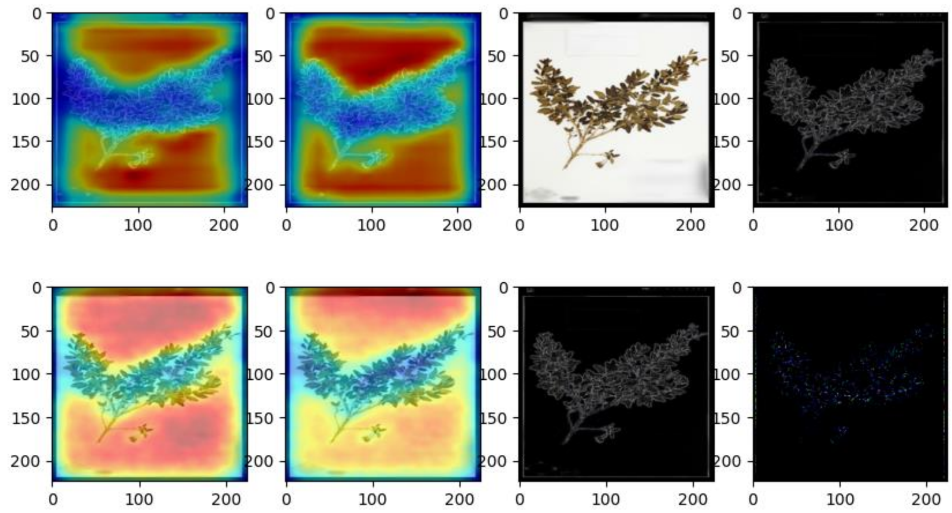
13- Find Edges Filter

Number of misclassifications: 46

Number of misclassifications after adding noise: 38

The application of the Find Edges filter was found to increase the misclassification rate of a CNN. This filter works by highlighting the edges and boundaries between objects in an image, which can introduce additional noise and make it more difficult for the CNN to correctly classify the image.

However, when adversarial noise was added to the images that had been filtered with the Find Edges filter, the misclassification rate decreased. This is likely because the adversarial noise helped to smooth out some of the noise introduced by the Find Edges filter, making it easier for the CNN to correctly classify the image.



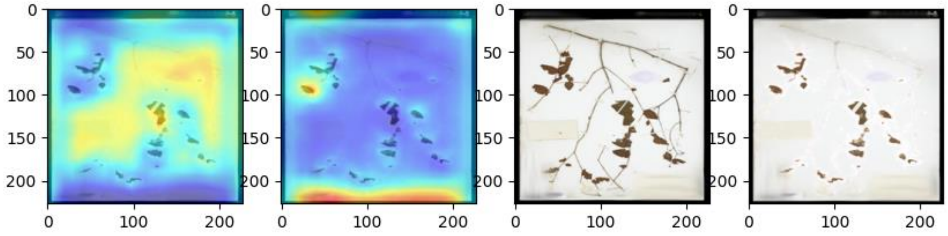
14- MaxFilter

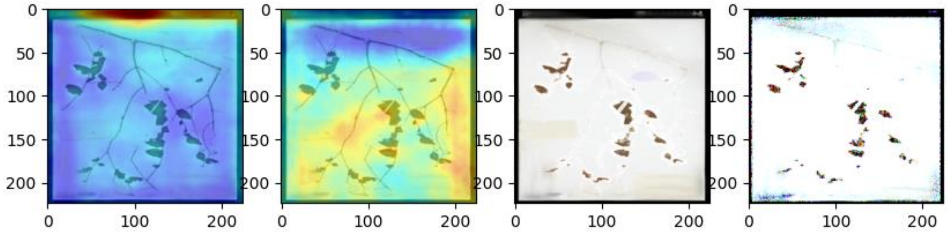
Number of misclassifications: 33

Number of misclassifications after adding noise: 46

The MaxFilter, which works by replacing each pixel in an image with the maximum value of the neighboring pixels, was found to increase the misclassification rate of a CNN.

Interestingly, when adversarial noise was added to the images that had been filtered with the MaxFilter, the misclassification rate increased even further. This suggests that the MaxFilter may introduce additional noise or artifacts into the image, making it more difficult for the CNN to correctly classify the image.





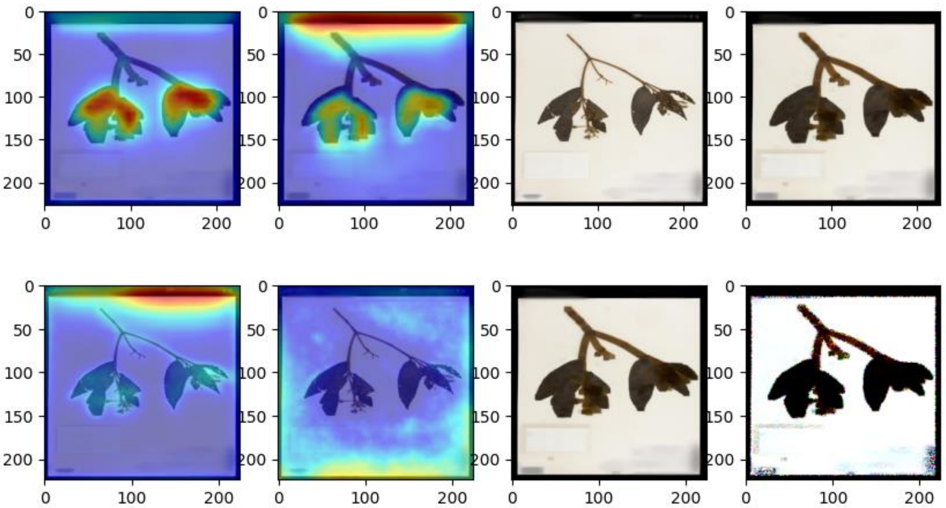
15- Min filter

Number of misclassifications: 45

Number of misclassifications after adding noise: 38

The min filter is a type of filter that replaces each pixel in an image with the minimum value of the pixels in its surrounding area. This filter is commonly used to reduce noise and smooth out the image. However, when this filter is applied to an image, it can also have the unintended effect of distorting the image in a way that makes it harder for the CNN to correctly classify.

After adding adversarial noise to the min filtered images, the misclassification rate decreased, meaning that the CNN was better able to correctly classify the images. This could be because the adversarial noise helped to counteract the distortions introduced by the min filter and made the images more recognizable to the CNN.


0
50
100
150
200
0
100
200