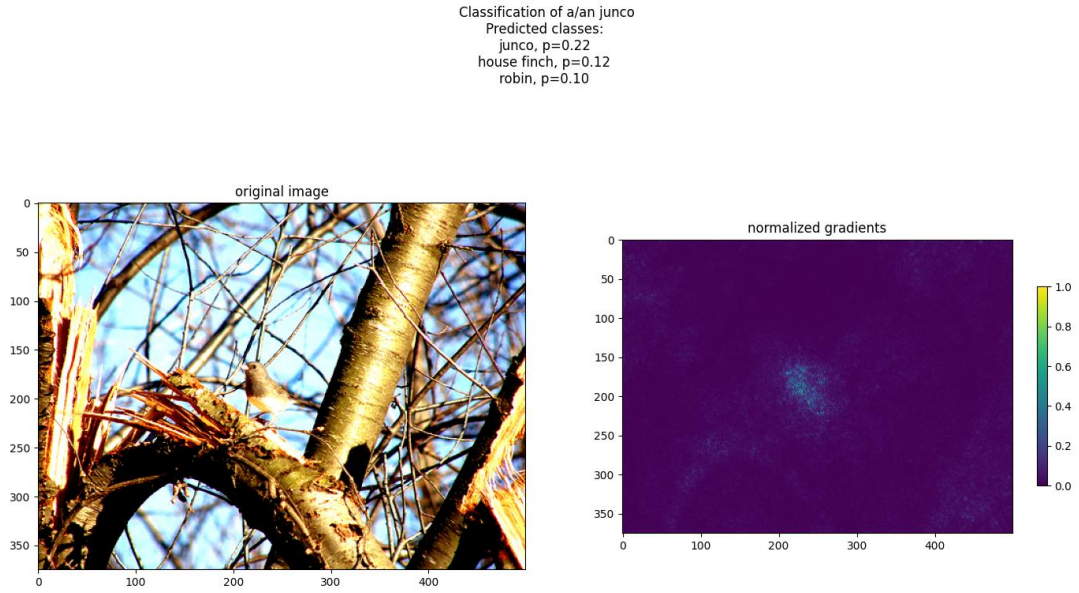
DEEP LEARNING EX 8

Uralp Ergin 5975013, İsmail Karabaş 7654321, Ahmet Yaşar Ayfer 5986167

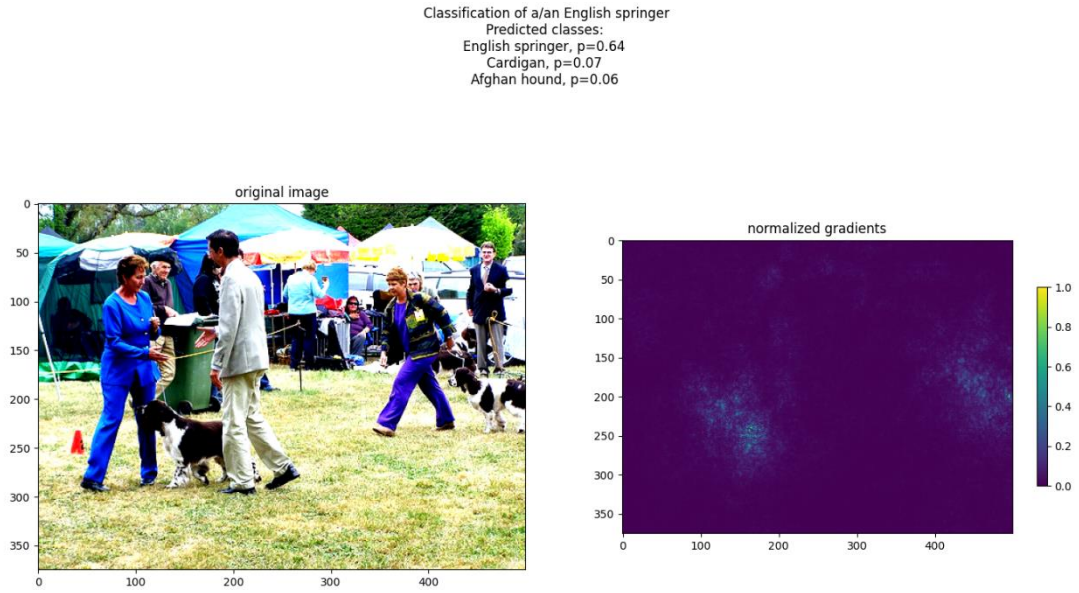
Attention On Input:

Figure1)



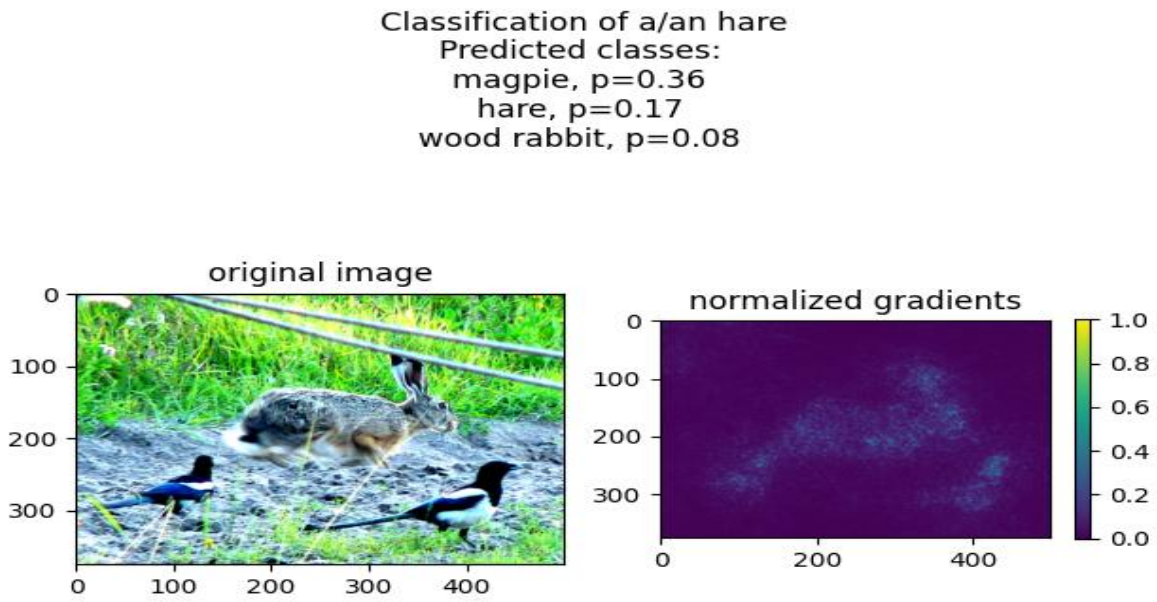
The gradients are higher at the middle part and lower at all other places. This is beacuse middle is the place where the bird stands.

Figure2)



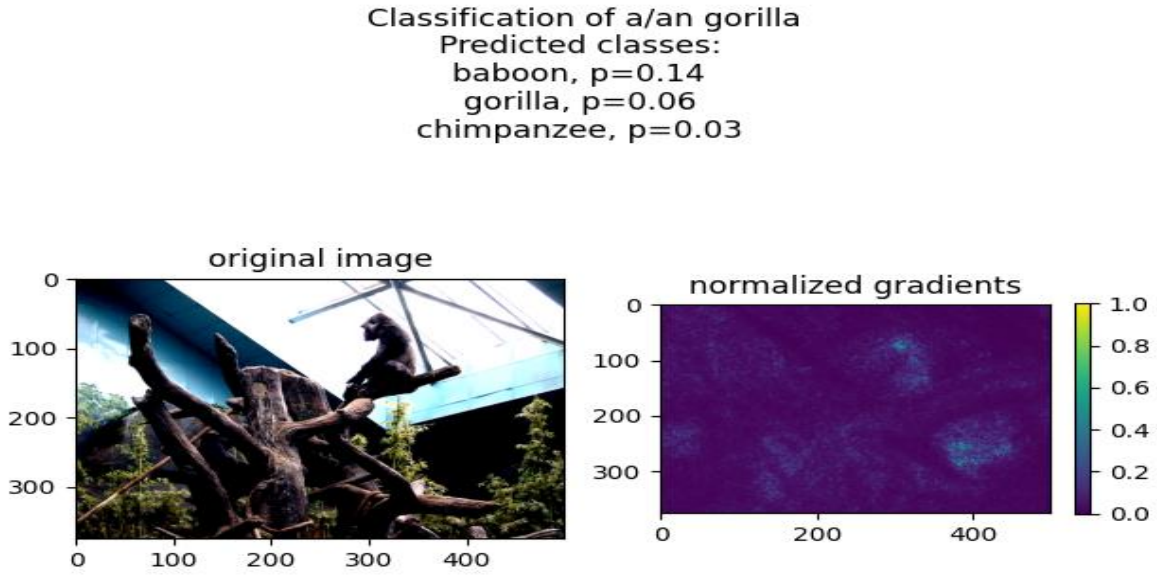
The gradients are higher at the left and right side and lower at all other places. This is beacuse these are the places where the dogs stands.

Figure3)



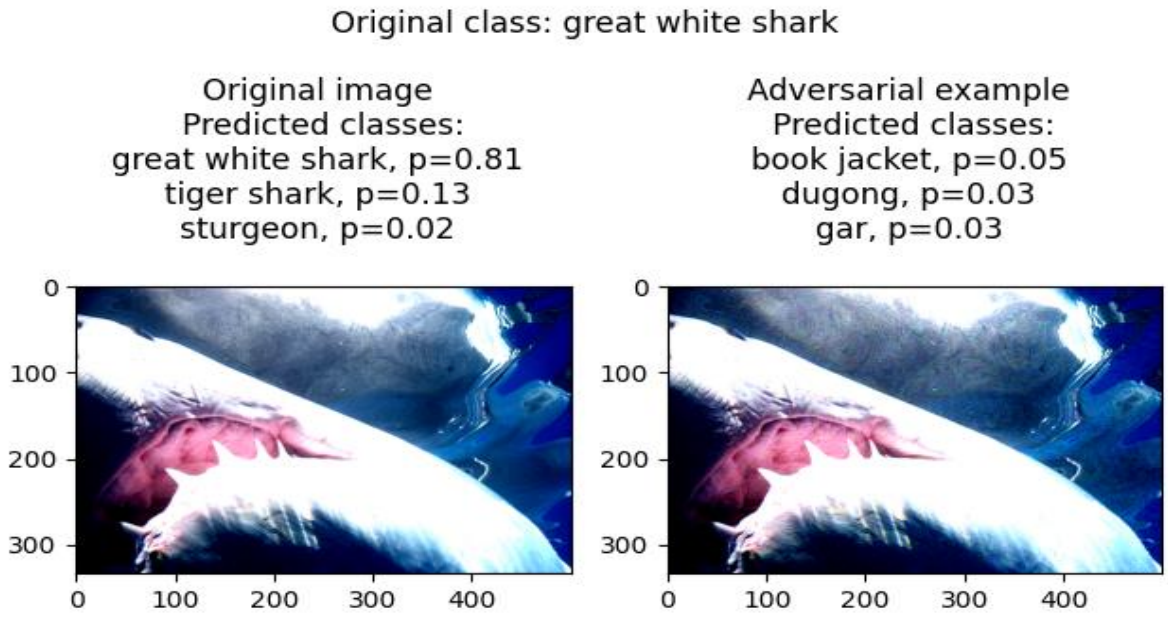
There are 3 places with high gradients at where 2 birds and 1 hare stands. The task is the classification of an hare but the model predicts magpie with higher probability which might be due to the similarities of their features that results in both high gradients.

Figure4)



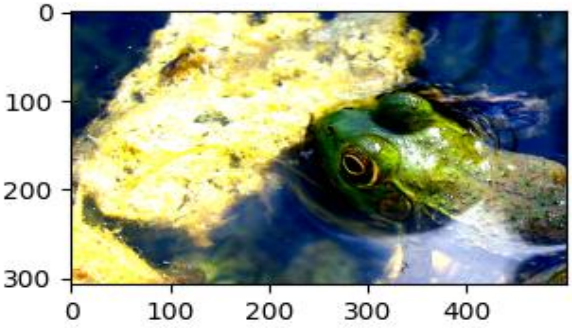
The normalized gradients show that the model focuses on regions around the gorilla, but it also highlights other areas, such as parts of the enclosure or branches. This behavior suggests that the model might not have fully learned the key features of gorillas, instead being distracted by irrelevant elements in the image.

Adversarial Examples:

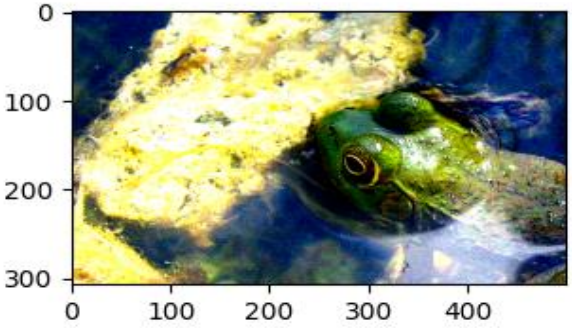


Original class: bullfrog

Original image
Predicted classes:
bullfrog, p=0.82
water snake, p=0.05
tailed frog, p=0.03

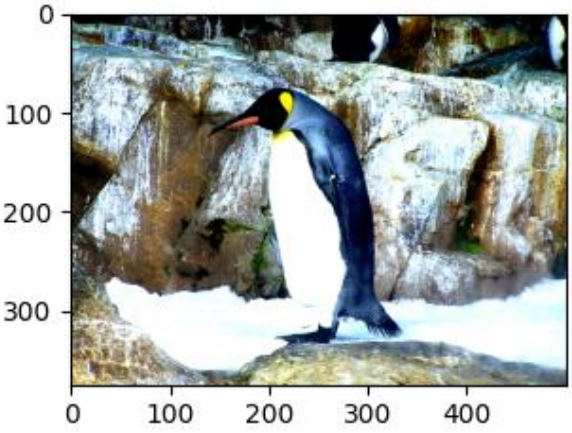


Adversarial example
Predicted classes:
jellyfish, p=0.17
coral reef, p=0.16
lionfish, p=0.06

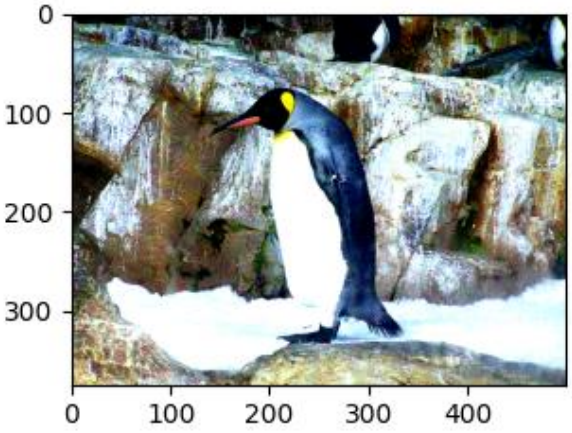


Original class: king penguin

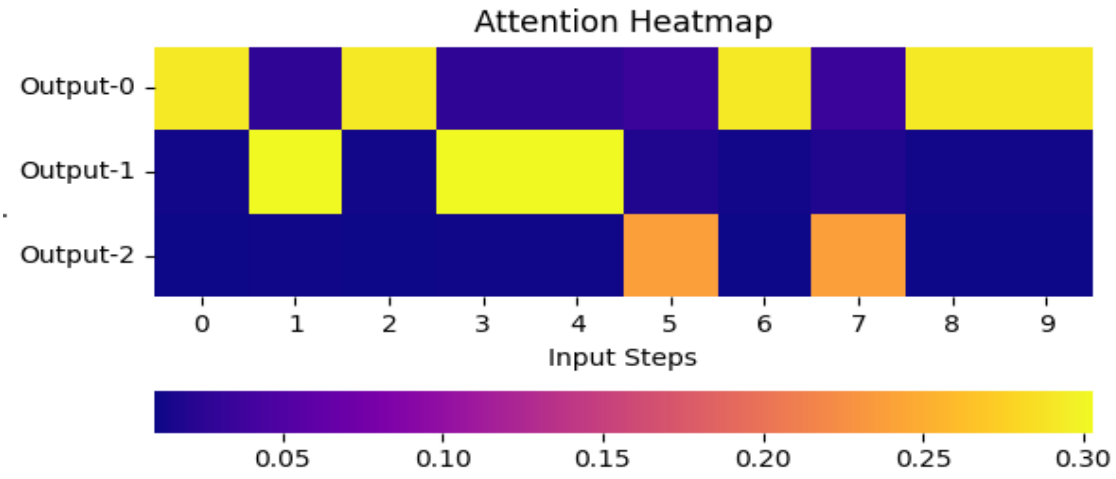
Original image
Predicted classes:
king penguin, p=0.98
oystercatcher, p=0.00
albatross, p=0.00



Adversarial example
Predicted classes:
American alligator, p=0.16
jigsaw puzzle, p=0.04
African crocodile, p=0.03



Counting Attention Plot:

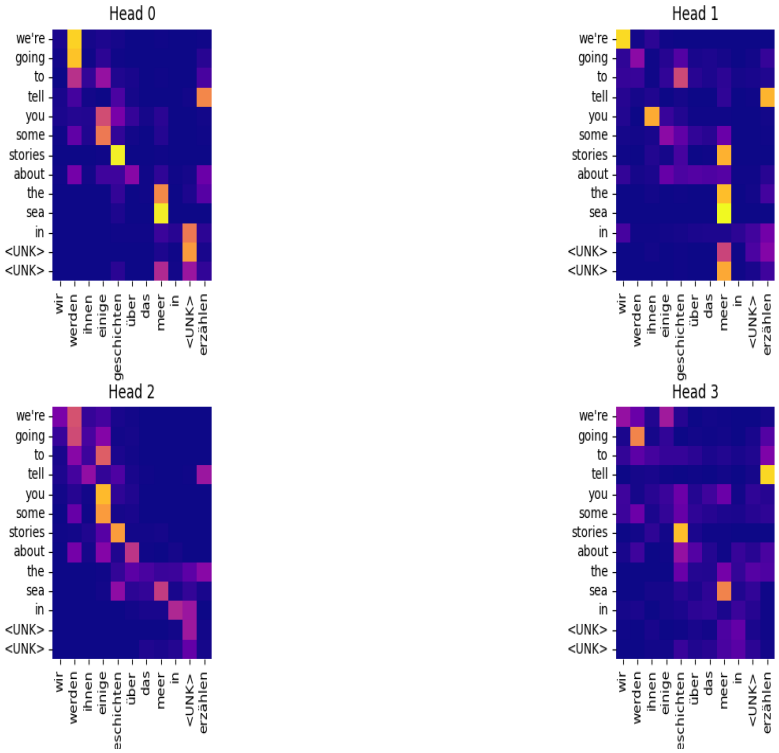


As it can be seen from the heatmap Output 0 attended the parts where there are "0" in the input with high attention (yellow). Output1 attended the parts where there are "1" in the input with high attention (yellow). Output 2 attended the parts where there are "2" in the input with 0.20 attention (orange).

(Bonus) Comparing Softmax and Sigmoid Usage for Attention:

Since Softmax normalizes the scores such that they sum to one, if you increase one score then you have to decrease others to preserve the sum. However, the outputs are independent and don't need tos um to one. Therefore, using Sigmoid which squashes all the outputs into the range 0-1 independently. This behaviour allows lots of the outputs to be high (give more attention).

Translation:



As it can be seen from the figures, different heads give different results which can be interpreted as they attend different parts (context) of the sentence.