Where's Whale-do Visualization

Ammar Ali

1 Theory for Visualization

The visualization of the extracted features from a deep neural network is so important to make sure that the model is on the track and not over-fitted. In addition, it gives clear insights about the behaviour of the model. For classification/feature extraction models, heatmaps are usually used for visualization based on a techniques called Gradient Class Activation Map (GRAD-CAM). This method weights the output of the last layer with the gradient taken for a certain class (the class we want to visualize in case of a classification problem). To apply this on a feature extractor, we can take the gradient for the summation of the feature values of the last layer in our case. To summarize the steps of the used method:

1. Get the model output and the last convolutional layer output for each image.
2. Get the sum of the embedding vector.
3. Compute the gradient of the summation with respect to the last convolutional layer.
4. Average and weight it with the last convolutional layer.
5. Normalize
6. Convert to RGB image.

1.1 Why Heatmap

Well almost all other visualization methods could be conducted from the heatmaps. For example, bounding boxes could be easily extracted from heatmaps by thresholding the resulting images we can get masks for the most important features / less important and so on. Then using these masks we can get the bounding boxes. We can use these highlighting ROIs with a matcher between two images to see the point to point relation. But in comparison between heatmaps and (bboxes, point to point matching), heatmaps have the advantage to show the priority of regions in a particular image. So, we can track the most important area. for example, if the whale has a scar, then this should be probably the most important feature to track it, then if there is no scare the most important

feature is dorsal fin and so on. and from that we can conduct if our model is behaving as it should be or not.

2 Examples

As a starting point, let us take a look at the features extracted from a random image taken from the training set for each model.

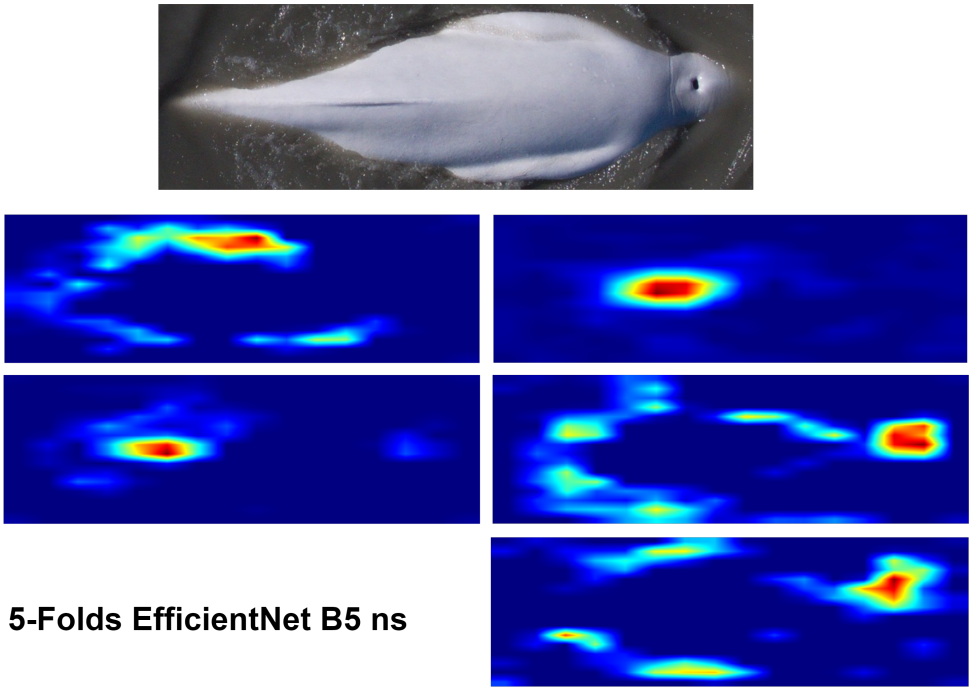

Figure 1: As we can see that these 5 folds highlighting the (dorsal fin, blowhole and edge shapes) as main parts

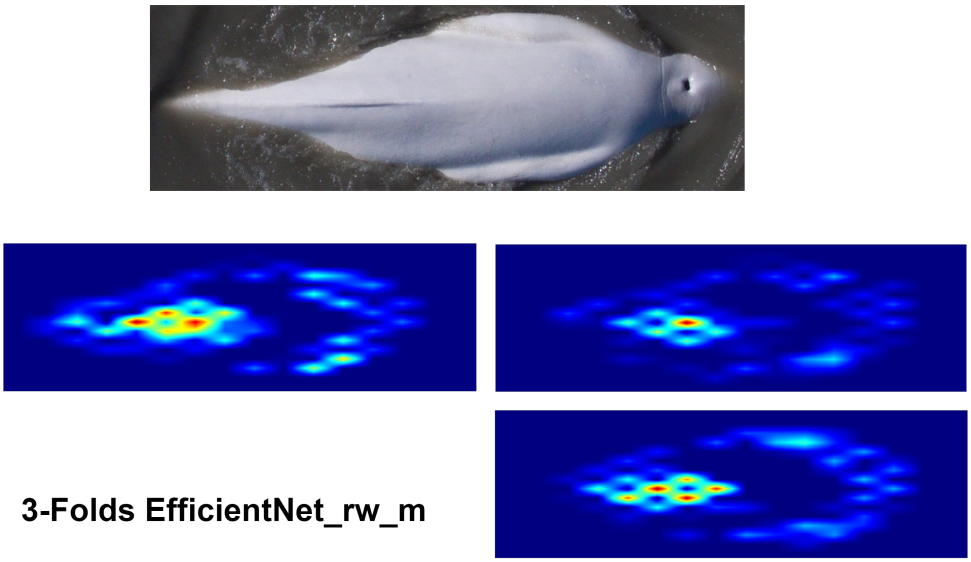

Figure 2: unlike that, EfficientNet-rw-m models are extracting similar features for this particular image from the dorsal fin part and around it

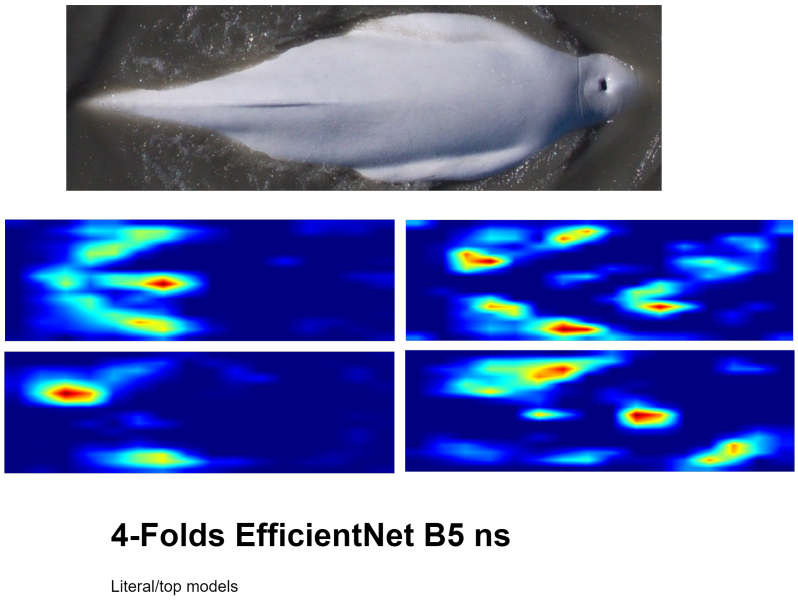

Figure 3: On the other hand EfficientNet b5 top/literal models are extracting features from the edges and sides of the whale to be able to compare with features for literal view images

2.1 top/top models

Now let us go to the interesting part and visualize the Union of the heatmaps with a basic thresholding and layer it to our image with the best match found for it.

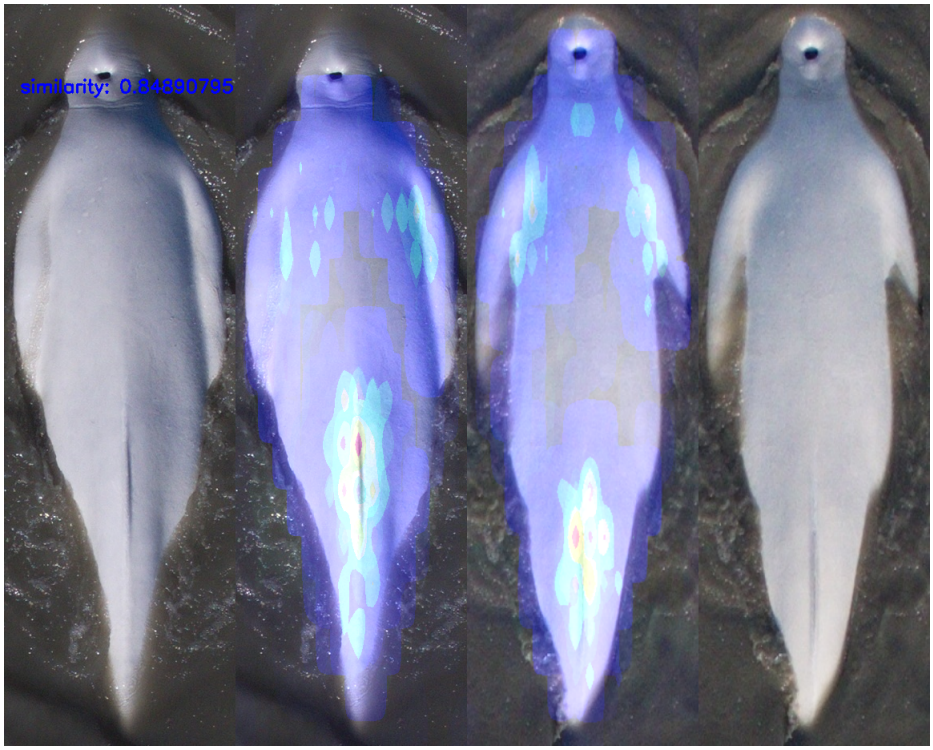

Figure 4: Left image is the query image, second image is the query image layered with the heatmap, third image is the top1 match layered with its heatmap and the right image is the top1 match image

As shown in fig(4), the extracted features almost segment the whale part, which is a really good indicator that the models are performing well. It is taking into consideration each pixel of the whale body viewed. which indicates that it will include the whale (color, illumination) into account but will not be considered as main features. and the main feature here are firstly the dorsal fin zone then the side edges. Let us take a look for another example where the whale has a scar on his body shown.

Figure 5: Left image is the query image, second image is the query image layered with the heatmap, third image is the top1 match layered with its heatmap and the right image is the top1 match image

In fig(5) , it is really clear how the models are prioritizing the scar features over dorsal fin especially in the matched image where the fin is really clear but the models are centered around the scar. As well the features are covering the full whale part which is again is a good indicator.

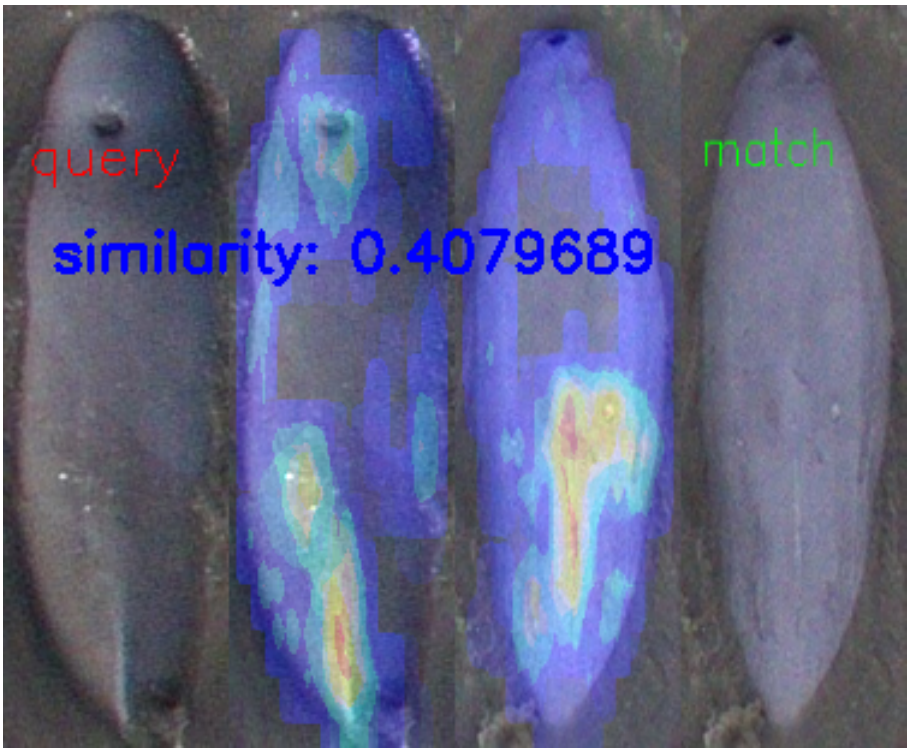

Figure 6: Left image is the query image, second image is the query image layered with the heatmap, third image is the top1 match layered with its heatmap and the right image is the top1 match image

In fig(6), we can see that the query image doesn't include a clear fin so the models also extracted features from the blowhole while on the best match the fin is clear so the main features are extracted from the fin and that is cause a low similarity in comparison with previous images which is logical.

2.2 top/literal models

Now let us take a look on how the top literal models will extract feature

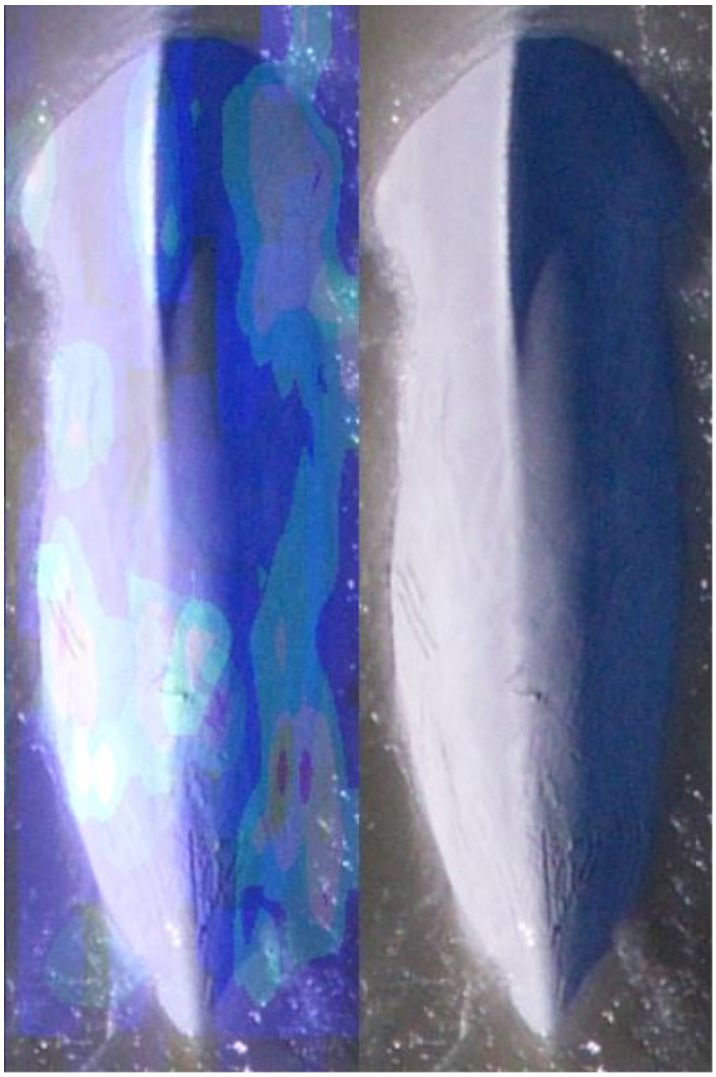

Figure 7: Left is the image layered with the heatmap, right is the original image

As shown in fig(7), the models are extracting the features from the sides of the whale, which is good because it needs to match these features with features of an image taken from the side literal view. On the other hand it has some noise (some features are taken from the water) because of the small amount of training images for literal view. and for sure this will be a problem and it is highly recommended to retrain the literal/top models when getting more data for these cases.

3 Observations

From the data visualization, we can conduct that the top/top models are performing well. actually while training firstly i trained the models validating again totally new whales to adjust parameters/thresholds/loss functions/ architecture. and these top/top models can get above 85% top1 and above 65% top20 for a single fold. but that is not reflected on the LB. even though i think it will be clear when testing over more and more data. We can conclude that the top/top models will work as the following:

1. If the whale has a scar, the model will match it with the whales that have a similar scar
2. If the whale doesn't have a scar but the dorsal fin is clear on the image, the models will match with whales that have similar dorsal find/ edges.
3. If none of the above, the models will match according to the blowhole shape
4. For all cases, the whale color and illumination will be taken into consideration but not as main features
5. The models are robust a against the background changes.
6. The main problem is that if a whale has a scar viewed on one image (other feature are not viewed) and not viewed on another image. In this case, the similarity will be low.

On the other hand, the top/literal models are not robust and have a lot of noise and a little over-fitting. even-though it does extract features from the whale sides but it also extract feature from the water background (indicating over-fitting) it is better to use these model to find only the best match then use top models to do the rest. Again it is highly recommended to retrain these models after gathering more data.