PadhAI: Visualising CNNs

One Fourth Labs

Occlusion experiments

Which patches in the image contribute maximally to the output?

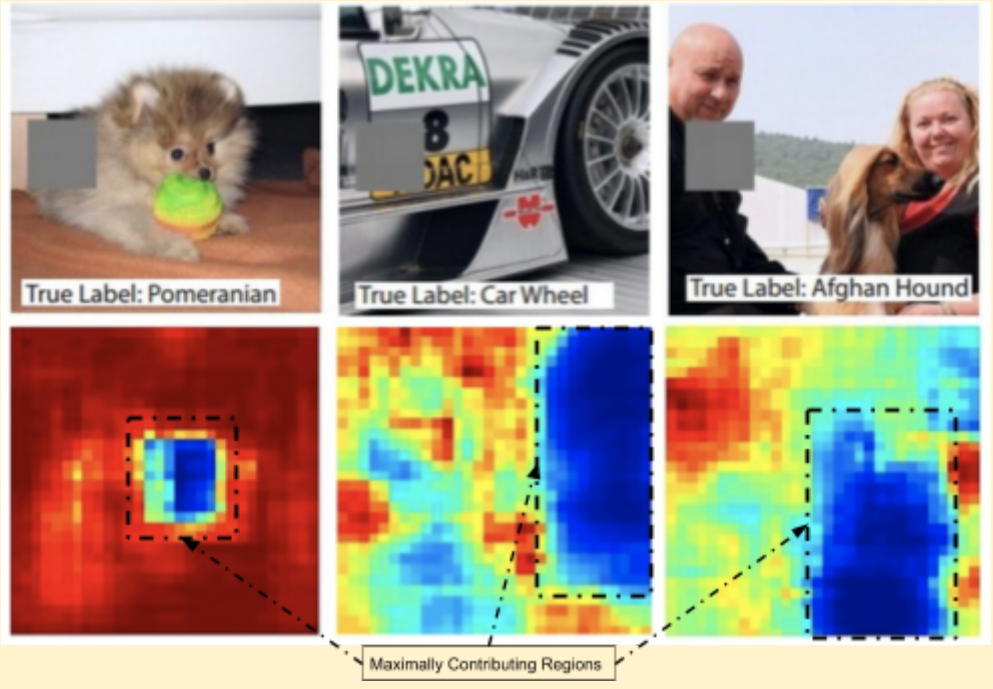
1. We can determine the maximally contributing regions in an image using occlusion experiments
2. Occlusion experiments refer to applying occlusions to the input image and seeing how they alter the predicted output distribution.
3. Then we take the difference between the un-occluded distribution and occluded distributions for occlusions placed in all patches of the images.
4. These differences allow us to generate a heatmap of maximally contributing patches.
5. Consider the following images with occlusions applied to them

   a. The true distribution for any of these images would be say $y = [... 1 ...]$
   b. The un-occluded predicted distribution $\hat{y}_1 = [.... 0.89 ....]$
   c. Occlusion placed at position 1 $\hat{y}_2 = [.... 0.84 ....]$

   ...

   d. Occlusion placed at position L $\hat{y}_n = [.... 0.3 ....]$

6. Now, calculating the difference between $(\hat{y}_1 and \sum_{n=2}^{L} y_n)$, we get differences in probilites which can be plotted as a heatmap over the input image with colder regions being more highly contributing.



---

7. Thus, the **higher the difference between un-occluded and occluded distributions**, the **more significant the contribution of that particular occluded region** to the output.