ES 654: Project Report

Vision with Lost Glasses: How does the Visual System deal with Noisy Conditions?

Abhigyan Martin Ninama
abhigyan.mn@iitgn.ac.in
IIT Gandhinagar

Aditya Shekhar
aditya.ss@iitgn.ac.in
IIT Gandhinagar

Anurag Kurle
kurle.anurag@iitgn.ac.in
IIT Gandhinagar

Ninad Shah
ninad.ps@iitgn.ac.in
IIT Gandhinagar

Rishabh Gupta
rishabh.g@iitgn.ac.in
IIT Gandhinagar

26 April, 2022

1 Abstract

Imagine you lost your spectacles and the world around you is completely blurred out. As you stumble around, you see a small animal walk towards you. Can you figure out what it is? Probably yes! In this situation, or in foggy/night-time conditions, visual input is of poor quality; images are blurred and have low contrast and yet our brains manage to recognize it. But how good are the current Deep Learning Models with these blurred out inputs? Deep Convolutional Neural Networks (DCNNs) are considered to be a good model of the ventral visual stream for visual classification. This project explores the performance of DCNNs on visual inputs with noisy conditions.

2 Introduction

Deep Convolutional Neural Networks (CNN or DCNN) are the type of neural networks that are highly efficient in problems related to object recognition by identifying patterns in images.

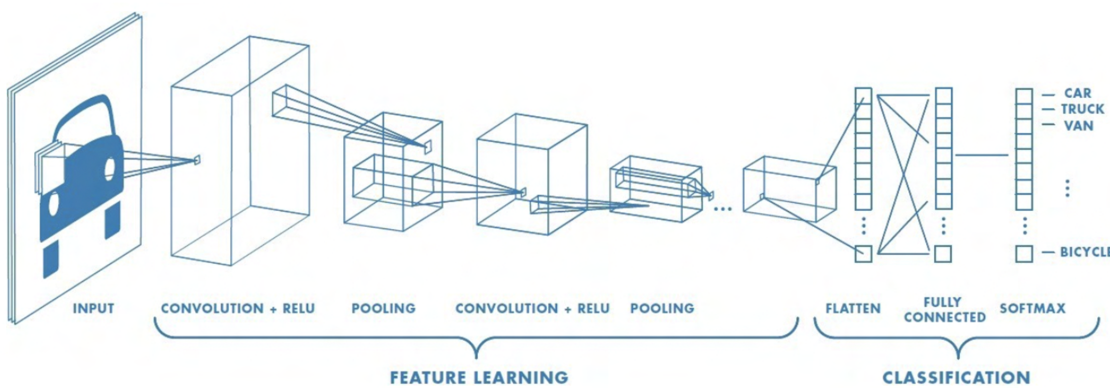


Figure 1: A Deep Convolutional Neural Network

To understand the similarity of deep convolutional neural networks to the human brain in the context of visual object recognition, researchers have defined a metric to measure which DCNNs are more brain-like, called brain-score. Higher brain-scores show higher brain-like performance of DCNNs for object recognition. Note that higher brain-scores do not ensure higher accuracies on the test dataset.

Brain-Score	model	neural predictivity V4	neural predictivity IT	behavioral predictivity	top-1 accuracy ImageNet
.549	densenet-169	.663	**.606**	.378	75.90
.544	cornet_s	.650	.600	.382	74.70
.542	resnet-101_v2	.653	.585	**.389**	77.00
.541	densenet-201	.655	.601	.368	77.00
.541	densenet-121	.657	.597	.369	74.50
.541	resnet-152_v2	.658	.589	.377	77.80
.540	resnet-50_v2	.653	.589	.377	75.60
.533	xception	.671	.565	.361	79.00
.532	inception_v2	.646	.593	.357	73.90
.532	inception_v1	.649	.583	.362	69.80
.531	resnet-18	.645	.583	.364	69.76
.530	nasnet_mobile	.650	.598	.342	74.00
.528	pnasnet_large	.644	.590	.351	**82.90**
.528	inception_resnet_v2	.639	.593	.352	80.40
.527	nasnet_large	.650	.591	.339	82.70
.527	best mobilenet	.613	.590	.377	69.80
.525	vgg-19	**.672**	.566	.338	71.10
.524	inception_v4	.628	.575	.371	80.20
.523	inception_v3	.646	.587	.335	78.00
.522	resnet-34	.629	.559	.378	73.30
.521	vgg-16	.669	.572	.321	71.50
.500	best basenet	.652	.592	.256	47.64
.488	alexnet	.631	.589	.245	57.70
.469	squeezenet1_1	.652	.553	.201	57.50
.454	squeezenet1_0	.641	.542	.180	57.50

Figure 2: Models and their respective brain scores.

To understand how these models deal with noisy visual inputs, some models will be selected from Figure 2 (Selection of Models) for further analysis.

Now, in order to explore the responses to these "black-box" neural networks on our noisy input, an extensive dataset has been created comprising of different types of noises. To implement these noises, we applied gaussian and cylindrical blurs on the normal dataset. These blurs were implemented on the images and the subsequent datasets contain low-blurred, medium-blurred and highly-blurred versions of these images. We have attempted at finding the correlation between the brain scores and accuracies by observing the trends in output of the models and analysing individual layer activations using GradCAM.

3 Simulating Real Eye Blur

In vision science, defocus is simulated in various ways, but by far the most common approach is to convolve parts of the scene with 2D Gaussians (Mather Smith, 2002; Watson Ahumada, 2011; Duchowski et al., 2014; Subedar Karam, 2016). In computer graphics, ray tracing is used to create depth-dependent blur in complex scenes (Cook, Porter, Carpenter, 1984). For non depth-varying scenes, this is equivalent to convolving the scene with a cylinder. But, for the accurate way to mimic real human eye blur, we refer to a paper on creating correct blur and its effect on accommodation(Steven A. Cholewaik, Gordon D. Love, Martin S. Banks, 2018).

The accurate way to mimic is by first calculating the retinal images by using the point spread function (PSF). According to wave optics for incoherent imaging (Goodman, 1996), the PSF at one wavelength,

$$PSF_{\lambda}(\theta, \phi)$$

is proportional to the square modulus of the Fourier transform of the complex aperture function, which takes into account both the amplitude and phase of the input:

$$PSF_{\lambda}(\theta, \phi) \propto \left| \mathcal{F}(Ae^{\frac{2\pi i}{\lambda}(Z_d + Z_{HOA} + Z_{LCA})} \right|^2$$

After this we calculate the displayed image by deconvolution of the input image, PSF display, and PSF infocus.

$$I_{retina} = (I_{input} \circledast PSF_{display}) \circledast PSF_{infocus}$$

Combining the equations, rearranging them and then taking fourier transform we get the following equation:

$$OTF_{display} = \frac{OTF_{defocus}}{OTF_{infocus}}$$

Here, OTF is the optical transfer function. With this we can define the kernel function for the calculated blur. But, this function is very complex with multiple fourier transformations as well as convolutions, if we consider the depth of image in mind, which can be hard to replicate. But, if we consider the image to be a 2D image then the complexity of this function reduces and it is very similar to a cylindrical blur and gaussian blur. This can be realized from the Figure 4.

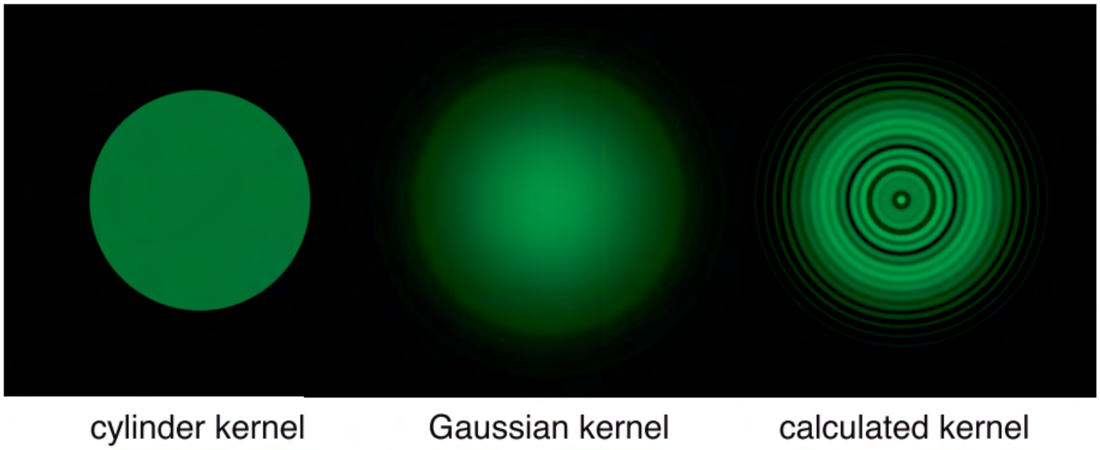


Figure 3: Blur Kernel comparision

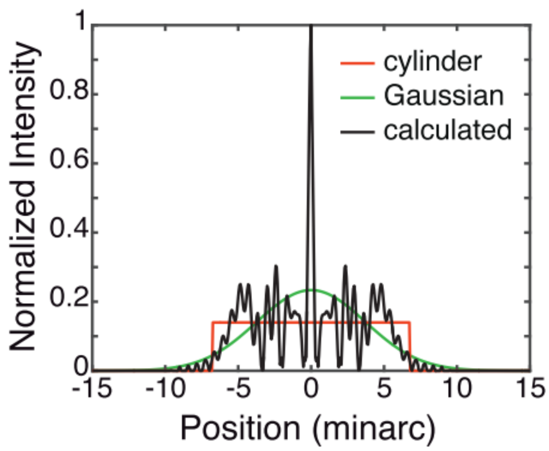


Figure 4: Position of image on Retina

Cylindrical and gaussian blurs are the closest approximations of the above transfer function in a 2D image. Hence, this project uses the blur kernels to be gaussian blur and cylindrical blur. These kernels would be used further in this project for creating datasets.

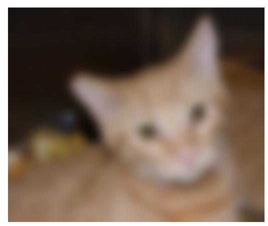
4 Datasets

Datasets play an important role in this project because they are the basis for classifying different models and calculate their accuracy. Our datasets include sharp and progressively blurred images that can be used to test different models and measure their accuracy. ImageNet is the base dataset on which all our models are pre-trained.

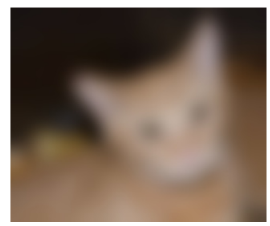
A new dataset was created with 2 classes (cats and dogs) with 2500 images of each class for training as well as testing. Next, multiple datasets were created with the same images while controlling the intensity of blur with cylindrical and gaussian blur. The sigma values for the gaussian blur were varied from 1 to 11 (**1,2,3,4,6,9,11**), whereas the kernel width of cylindrical blur was varied from 5 to 17 (**5,9,13,17**). This made the overall dataset to contain as many as 120,000 images.

(a) Original Image with no blur

(b) Image with Gaussian Blur($\sigma = 1$)

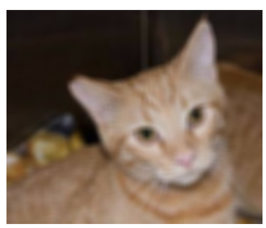


(c) Image with Gaussian Blur($\sigma = 6$)

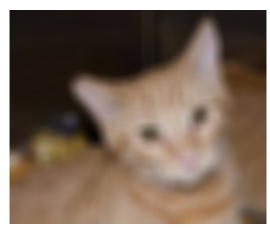


(d) Image with Gaussian Blur($\sigma = 11$)

(e) Image with Cylindrical Blur(kernel width=5)



(f) Image with Cylindrical Blur(kernel width=9)



(g) Image with Cylindrical Blur(kernel width=17)

Figure 5: Our Dataset

5 Selection of Models

Selection of models is an important aspect of this project. Hence, the models have been selected (Figure 6) carefully by taking into consideration the following points:

Brain-Score	model	neural predictivity V4	neural predictivity IT	behavioral predictivity	top-1 accuracy ImageNet
.549	densenet-169	.663	**.606**	.378	75.90
.544	cornet_s	.650	.600	.382	74.70
.542	resnet-101_v2	.653	.585	**.389**	77.00
.541	densenet-201	.655	.601	.368	77.00
.541	densenet-121	.657	.597	.369	74.50
.541	resnet-152_v2	.658	.589	.377	77.80
.540	resnet-50_v2	.653	.589	.377	75.60
.533	xception	.671	.565	.361	79.00
.532	inception_v2	.646	.593	.357	73.90
.532	inception_v1	.649	.583	.362	69.80
.531	resnet-18	.645	.583	.364	69.76
.530	nasnet_mobile	.650	.598	.342	74.00
.528	pnasnet_large	.644	.590	.351	**82.90**
.528	inception_resnet_v2	.639	.593	.352	80.40
.527	nasnet_large	.650	.591	.339	82.70
.527	best mobilenet	.613	.590	.377	69.80
.525	vgg-19	**.672**	.566	.338	71.10
.524	inception_v4	.628	.575	.371	80.20
.523	inception_v3	.646	.587	.335	78.00
.522	resnet-34	.629	.559	.378	73.30
.521	vgg-16	.669	.572	.321	71.50
.500	best basenet	.652	.592	.256	47.64
.488	alexnet	.631	.589	.245	57.70
.469	squeezenet1_1	.652	.553	.201	57.50
.454	squeezenet1_0	.641	.542	.180	57.50

Figure 6: Selection of suitable models

To account for different brain scores with similar accuracy, **ResNet-50_V2** and **Inception_V3** have been selected. Both models have Top 1% accuracy values on ImageNet as 77.80 and 78.00 and have brain scores of 0.540 and 0.523 respectively. **Inception_ResNet_V2** has been chosen as the third model into consideration because it has high accuracy (80.40) and with an average value of brain score (0.528). Finally, **ResNet-152_V2** is chosen because it has similar brain score (0.541) to the above models but has a different values of accuracy.

The above selected models will be used for further analysis. These models will be responsible for providing the overview of the models with different brain scores and varying accuracy values.

6 Observations and Inferences

This section compares the accuracy of the selected models that were trained on normal datasets as well as blurred datasets. The models were trained using transfer learning with the pre-trained weights of the models on ImageNet Dataset. The final layer of the selected models were trained and tested against our dataset. The accuracies of the models were noted as we kept increasing the blur intensity.

6.1 Accuracy Curves and their Analysis

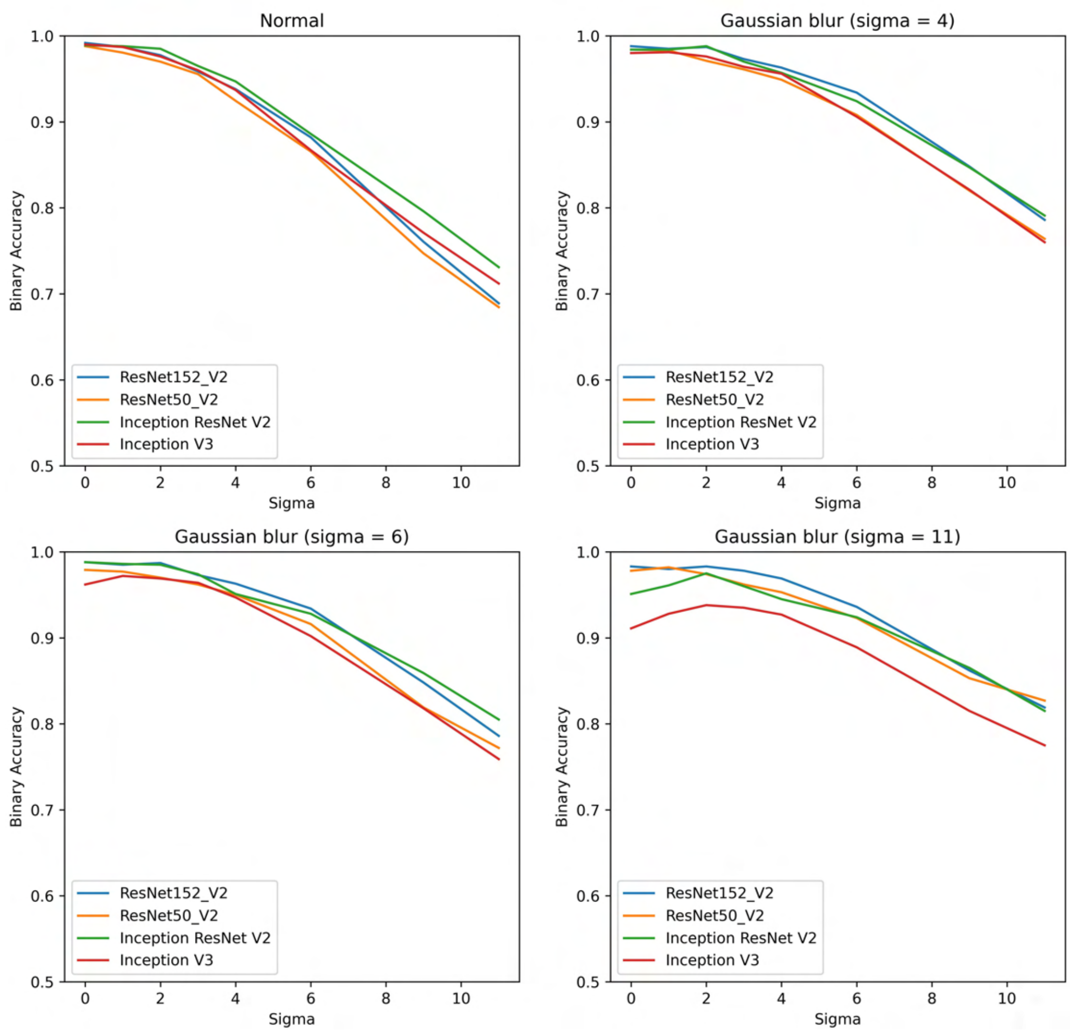


Figure 7: Performance on Gaussian Dataset

We have plotted the graphs for the selected models trained on Gaussian datasets (sigma = 4,6,11) and tested on Gaussian datasets.From these graphs, we observe that as we train the models on blurred images the overall accuracy increase but every model behaves differently. We see that when the model is trained on normal dataset it performs very poorly while blur intensity is increased. While we train the models with different blurred datasets, we observe that overall accuracy does improve but we cannot differentiate much between them. But, it is quite clear that as the blur intensity is increased, the overall accuracy of **Inception_V3** is significantly lower than the other models. The next subsection does an in-depth analysis of the effects of varying blur intensity. The exact values of the accuracy found while training/testing the model can be found here: https://rb.gy/dqroj4

6.2 Model Performance Analysis

In the previous section, *Selection of Models*, we carefully selected the models in a manner such that we have different brain scores and different top 1% accuracy. From figure 6 we also have the values for Brain score, behaviour prediction and, ImageNet top-1 accuracy for each model. We used these indicators to see how they relate to a model's performance on normal dataset.

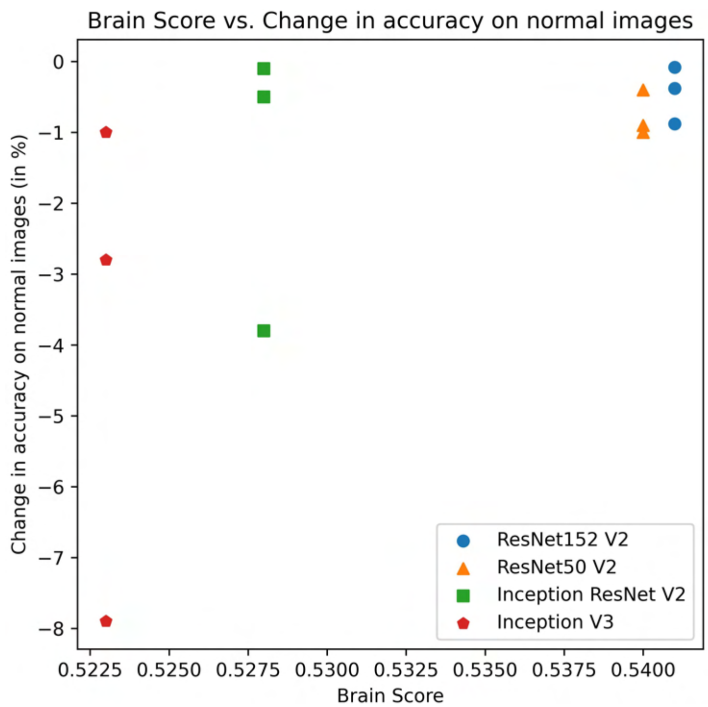


Figure 8: Brain Score vs. Change in accuracy on normal images

From this graph we see that the accuracy drop is strongly dependent on the brain score. As the brain score of the model increases we see that the drop in accuracy for normal dataset decreases as well. This means that the models with higher brain scores looks for the coarser details in the images to classify and hence their accuracy drop is low even with blurring the images.

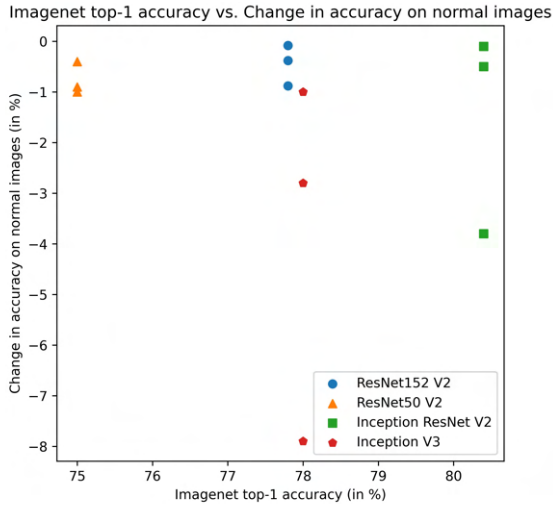


Figure 9: Behavior prediction rate vs. Change in accuracy on normal images

Since, the models that we have chosen are among the best, to account for the effect of high accuracy we also graphed the accuracy drop against the ImageNet top 1% accuracy. We can see that even though **Inception_V3** has a high top 1% accuracy but its drop in accuracy is very high. This suggests that the top 1% accuracy has very low correlation to how the model will perform with blurred images.

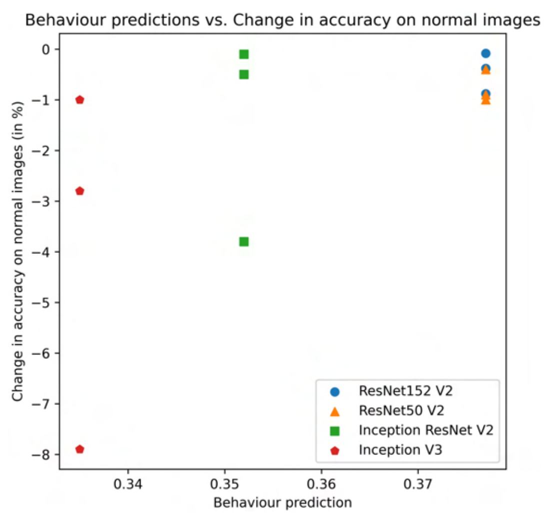


Figure 10: ImageNet top-1 accuracy vs. Change in accuracy on normal images

Here, it is also visible that as the behaviour prediction rate increases, the drop in accuracy decreases. So from these graphs, we found that higher brain scores and behaviour prediction rates correlated with lower drop in accuracy when trained on blurred dataset. This shows us that these models use coarser features to differentiate between cats and dogs. Thus, we can conclude that higher brain scores is a good indicator of how well a model can perform with blurred vision.

6.3 GradCAM Analysis

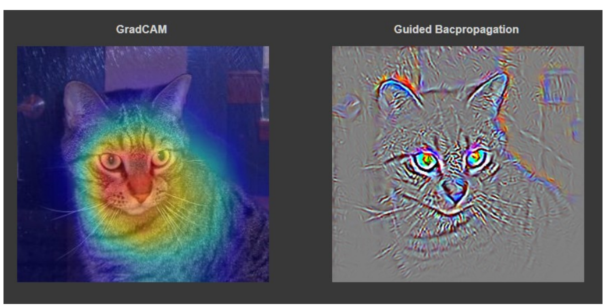
While deep learning has enabled exceptional accuracy in image classification, object recognition, and picture segmentation, one of its most significant issues is model interpretability, essential for model comprehension and debugging.

Neural networks (including CNNs and DCNNs) are often referred to as "black-boxes", that take in input and predicts the output; but it is often difficult to interpret *how* they reach the output prediction. Grad-CAM or gradient-weighted class activation maps is method that helps us answer this *how*. GradCAM assists in accurately comprehending where they're "seeing" in an image. This is done by creating a heatmap of the important parts of the images that were taken into consideration while training the last layer of the model.

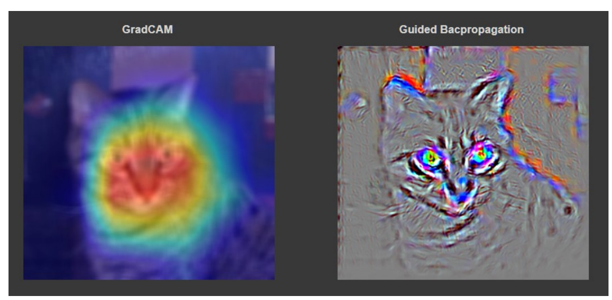
GradCAM has helped the practitioners in visually debugging their models. We can visually confirm where our network is looking with GradCAM, ensuring that it looks at the correct patterns in the picture and activates around them. Due to the above reasons, GradCAM is used in our study to learn how different models react to different intensities of blur.

As we can see the activations of the final convolution layer of ResNet-50_V2, on differently blurred images let us analyse the obtained GradCAM outputs.

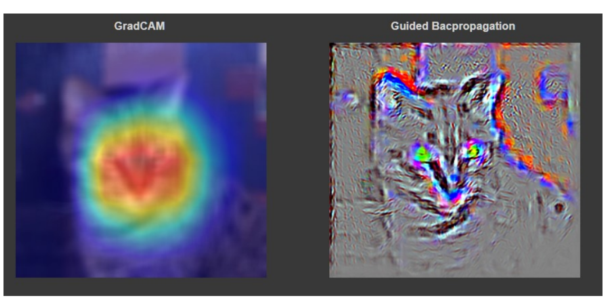
For each image we have the heat map of the model activations interfaced on the original image along with its guided back propagation showing the edges due to the model activations.



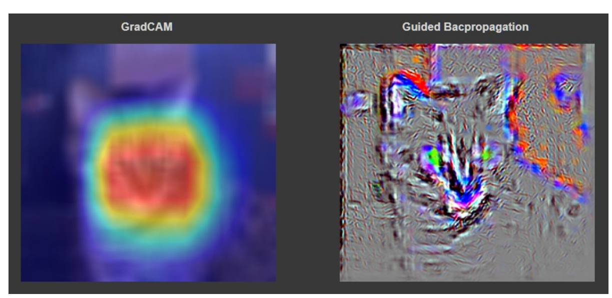
(a) Original Image with no blur



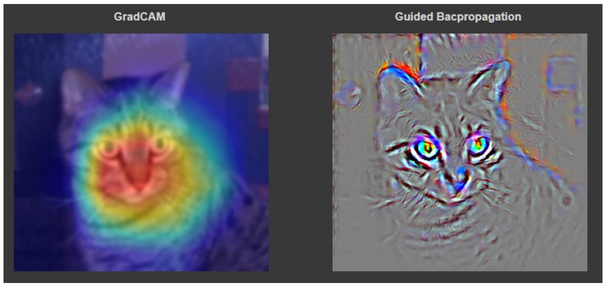
(b) Cylindrical Blur (kernel width = 9)



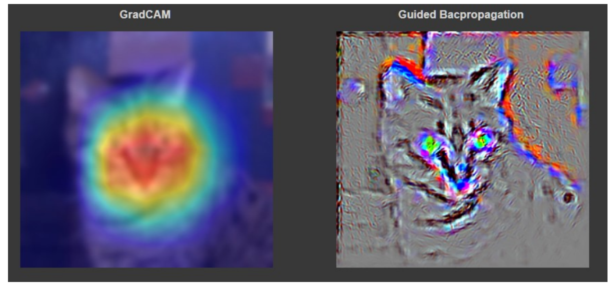
(c) Cylindrical Blur (kernel width = 13)



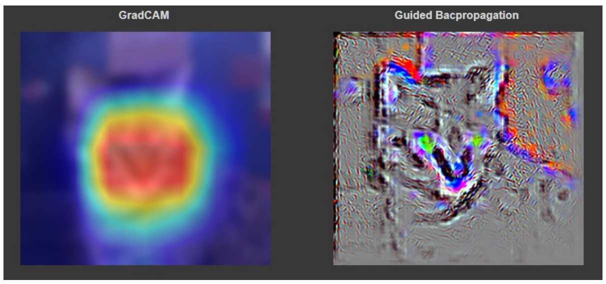
(d) Cylindrical Blur (kernel width = 17)



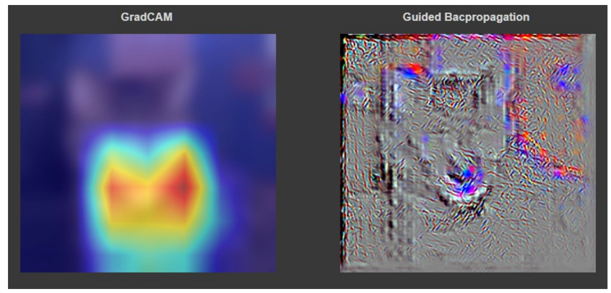
(a) Gaussian Blur ($\sigma = 2$)



(b) Gaussian Blur ($\sigma = 4$)



(c) Gaussian Blur ($\sigma = 6$)



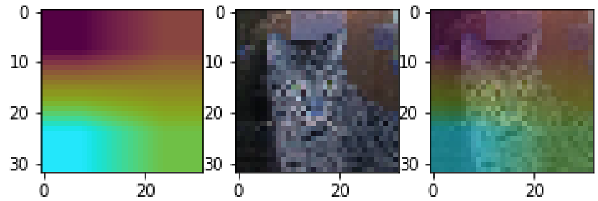
(d) Gaussian Blur ($\sigma = 9$)

Figure 12: GradCAM output on ResNet50V2 images

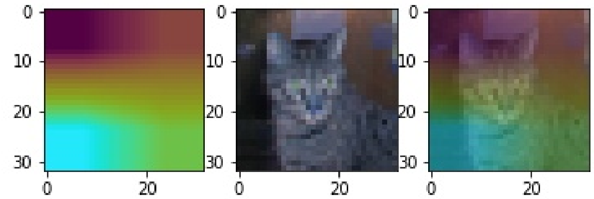
We can observe that in the images added with cylindrical blur, as we keep on increasing the blur strength, ResNet is able to detect the subject but the activation regions varies with the blur strengths. As we keep on increasing the blur strength, the activation region starts to shrink, or in other words, the model is finding it hard to find features corresponding to cat and it settles for more easily observable features such as cat whiskers or nose.

Now for the case of Gaussian blur, as we keep on increasing the blur strength, the activation region starts to shrink towards the more easily perceivable features of the subject as mentioned before. But as we reach to the point where the blur intensity is so high such that it is only able to detect the whiskers for it to be able to identify the subject as a cat. This phenomenon can also be confirmed through the guided back propagation in which we can see the gradual feature edges decrease.

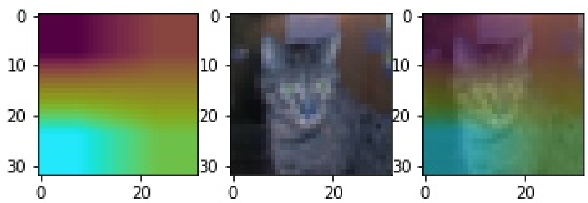
This same phenomenon can be seen in the InceptionV3 model which is expected since it is also a DCNN, but the trends in the change of feature activation varies from model to model which depends on the types of layers used in the specific model.



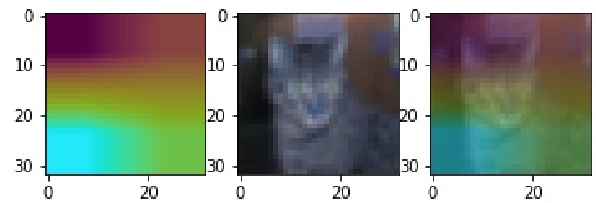
(a) Original Image with no blur



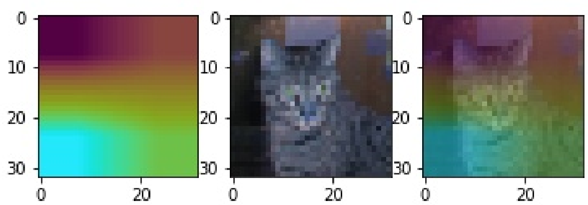
(b) Cylindrical Blur (kernel width = 9)



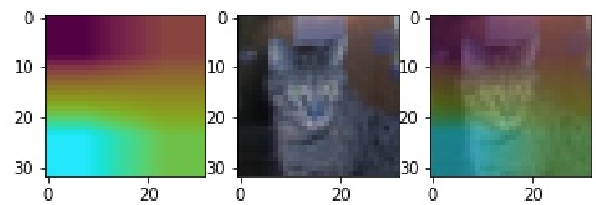
(c) Cylindrical Blur (kernel width = 13)



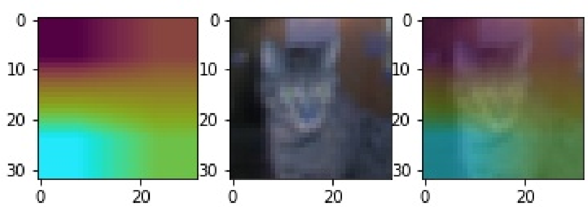
(d) Cylindrical Blur (kernel width = 17)



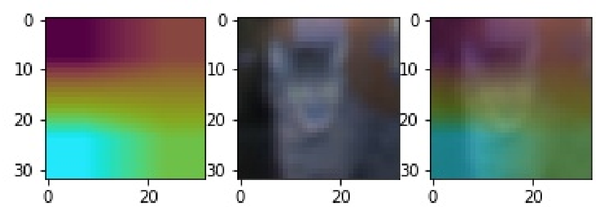
(e) Gaussian Blur ($\sigma = 2$)



(f) Gaussian Blur ($\sigma = 4$)



(g) Gaussian Blur ($\sigma = 6$)



(h) Gaussian Blur ($\sigma = 9$)

Figure 13: GradCAM output on InceptionV3 images

7 Conclusion

This brings us to the conclusion of this project. We conclude that models with high brain score perform relatively better when they are trained with blurred images. This shows that high behaviour prediction rate suggests that models are trained on coarser details and work better with blurred images. Although, models like **Inception_V3** (low brain score) have high top 1% accuracy, but if the dataset is blurred they work rather poorly. Whereas, models like **ResNet50_V2** (high brain score) have rather low top 1% accuracy, it will perform better than most model if the dataset is blurred.

7.1 Future Scope

We have found that there is a correlation between brain scores and learning ability of models on blurry/noisy data. Going forward, following avenues can be explored:

1. Studies can be done involving human participants to quantify at what level of detail the human brain works to differentiate between objects and compare the results with our models.

2. We can see how well the models with high brain scores perform to noisy data in other tasks such as imperfect information games.

3. We have approximated the real eye blur to gaussian/cylindrical blur for the scope of this project. However, we can get improved results by using better modelling of the real eye blur. Another possible improvement over our current results could be using ray traced blurring on images with depth and use of more sophisticated kernels. So that, the study can be performed on even more realistic data.

4. The problem can also be approached in another way, wherein we use deblur kernel to remove noise from the original images and then train our model using these newly curated dataset.

8 Acknowledgements

Given the extensive analysis required, this project would not have been possible all by ourselves. We would like to thank Prof. Nipun Batra for giving us an opportunity to work on this problem statement. We would also like to thank our mentor, Shriraj Sawant, for guiding us throughout the course of this project and providing inputs in the observations of our results, which allowed us to perform more trials in a shorter amount of time.

References

[1] Schrimpf, M., Kubilius, J., Hong, H., Majaj, N., Rajalingham, R., Issa, E., Kar, K., Bashivan, P., Prescott-Roy, J., Geiger, F., Schmidt, K., Yamins, D. and DiCarlo, J. *Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?*, 2022. URL https://www.biorxiv.org/content/10.1101/407007v1

[2] Billauer.co.il. *Blur simulator: What you'll see without your glasses (depending on prescription: sphere, cylinder and axis)*, 2022. URL http://billauer.co.il/simulator.html

[3] *CooperVision Contact Lenses | CooperVision UK*, 2022. (Online Visualisation) URL https://coopervision.co.uk/practitioner/clinical-resources/myopia-in-children/myopia-simulator

[4] Karen Simonyan, Andrea Vedaldi, Andrew Zisserman. *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*, 19th April 2014. URL https://arxiv.org/pdf/1312.6034.pdf

[5] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra | Georgia Institute of Technology. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization.*, URL https://openaccess.thecvf.com/content_ICCV_2017/papers/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.pdf