Intuition Report 9

Ramtin Mojtahedi Saffari - 20307293

Assessed link 1: OPENAI Microscope

Self-reflection

- **Short Summary**

The ink provides a collection of representations of every important layer and neuron of numerous popular “model organisms,” which are often investigated for their interpretability. The proposed Microscope makes it simpler to evaluate the characteristics that arise within these neural networks, and we believe that it will benefit the scientific community as we seek to comprehend these complex systems.

- **Hypothesis and Expectation**

I hypothesize that in a deeper network, the DeepDream visualisation of edges, textures, patterns, parts, and objects in images is distiguished and clearer than in a less complex network, which is visually more distinguishable within hidden layers.

- **What I Achieved and Learned**

Deep Dream is a deep learning feature visualization approach that synthesizes visuals with high network layer activation. The image characteristics learned by a network may be highlighted by displaying these images. Then, the images can be used to diagnose and analyze network activity [1]. I tested the proposed hypothesis using VGG19 and CLIP Reset-101. The former has 27 nodes, while the latter has 141 nodes and is much deeper. By assessing the provided feature visualizations through the DeepDream technique, I understand that the first convolutional layers learn edges and basic textures. Later convolutional layers pick up on details like textures and patterns that are more intricate. Objects or fragments of things are learned in the final convolutional layers. The fully connected layers learn to link the activations from high-level characteristics to the expected individual classes. By assessing hidden and intermediate layers where objects' visualizations are represented, the Resnet-101 provides a clearer and distinguishable visual representation compared to the hidden layers of VGG19. Both networks are using standard images in the Imagenet dataset. Therefore, the results can be comparable. Also, as we go through the last layers, such as the classifier layer, we can achieve that the Resnet provides a detailed and better representation of the image object parts, leading to better classification performance.

Through assessment of the reference [2], it is obvious that the deeper network of Resnet-101 performs better than VGG19 on the Imagenet dataset in classification tasks with about 1.08 times higher performance. However, its training time is approximately 1.24 times slower than VGG19. Considering the level of the feature representation in different nodes and performance, it still outperforms VGG19. The same pattern is also achieved by assessing large networks and smaller ones(e.g., VGG16).

Suggestion and Filling the Gaps

I found the interactive link very interesting and helpful in learning about the different techniques and functionality of feature visualizations in CNN networks. It is recommended to add in-detail short and publicly understandable explanations to the provided website, such as using information given in the reference [3].

Assessed link 2: Network Dissection

Self-reflection

- **Short Summary**

The link discusses the idea of nwrok dissection. Network dissection is characterized as interpreting networks by assigning meaningful labels to their constituent units, which indicates that deep representations are more transparent and interpretable than they would be if they were based on a random, equivalently powerful foundation.

- **Hypothesis**

I hypothesize that color is the lowest type of feature that can be represented in networks. While the scene is among the highest representable features, this trend exists among large and small networks. However, in self-supervised tasks, the texture is among the highest distinguishable features. Overall, larger networks work better than smaller networks in distinguishing and detecting features.

- **Testing the hypothesize and what I have learned**

Through assessment of Resnet-152 and VGG network, it is understandable that the number of unique nodes that could correctly detect scenes is about two times higher than the VGG when assessing the performance on the place365 dataset with

almost ten million images. When it comes to evaluating the same networks on imagenet with about 14 million imges, the resnet-152 detects scenes about four times higher than VGG. However, VGG shows higher detection on objects while Resnet-152 still detects scenes among the majority of its nodes. This demonstrates that the depth of the network is also one of the factors in a higher detection among all of the features and the same pattern existed between smaller and larger networks. Here, Resnet-152 consists of 50 nodes, while VGG consists of 27 nodes [4]. As discussed in the link, the form of tasks, whether supervised or self-supervised, is another factor involved in feature representation. For example, objects are among the highest distinguishability among supervised tasks in different networks, regardless of small or large. However, in self-supervised tasks, the texture becomes among the highest distinguishable feature. However, overall, objects due to their spatial features and the fact that they are being learnt within hidden layers are among the highest features that are being detected.

On another analysis, when the images are being transformed and changed compared, This experiment shows that the feature detection is being decreased among all feature types. However, the pattern remains with the highest detection for objects and lowest for color and material. This may also highlight the significance of data augmentation in a trianed network. The network can be used in a more generalized format and increase the performance of feature detection.

Suggestion and Filling the Gaps

Like I mentioned before, I found the interactive link very interesting and helpful in learning by visualization and graphs with different given graphs and experiments. It is recommended to provide more explanations for how the graphs and experiments have been achieved and conducted. This information can be used from the reference [5].

Self-evaluation:

In this intuition report, I have gone through an in-depth analysis of two of the provided links. In my assessment, I have completely considered the required expectation for deep exploration, including proposing a hypothesis and what I expected, reporting on what I achieved and explored, in-depth discussion of what I have learned, and providing gaps and recommendations to fill them. Considering the quality and assessment level, I deserve to get the full mark (4 points) for this intuition report.

In advance, thank you very much for your time and consideration of this report.

Best regards,

Ramtin

References

[1] 4.0, T. I. I. (2021, September 3). *An Introduction to DeepDream - Artificial Intelligence in Plain English*. Medium. https://ai.plainenglish.io/deepdream-ebe2a8016b29

[2] Team, K. (2021). *Keras documentation: Keras Applications*. Pre-Trained Networks. https://keras.io/api/applications/

[3] Mukherjee, S. (2020, January 26). *Do You Dream Of Me? A tutorial on DeepDream*. DEV Community. https://dev.to/subhadityamukherjee/do-you-dream-of-me-a-tutorial-on-deepdream-3mip

[4] Anwar, A. (2021, June 7). *Difference between AlexNet, VGGNet, ResNet, and Inception*. Medium. https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaaecccc96

[5] Zhou, B., Bau, D., Oliva, A., & Torralba, A. (2019). Interpreting Deep Visual Representations via Network Dissection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(9), 2131–2145. https://doi.org/10.1109/tpami.2018.2858759