What does Deep CNN learn?

Visualization of Popular Deep CNN Models

孫凡耕 羅啟心 許晉嘉 郭子生

National Taiwan University, Department of Electrical Engineering

Abstract

Deep Convolutional Neural Networks (CNN) revolutionized Computer Vision in recent years. However there is no direct understanding or derivation of why they perform so well, or how they might be improved. In our final project, we organize all common methods to visualize and understand CNN on the pretrained and popular AlexNet, VGG, GooLeNet and ResNet. This enables us to build up understanding of CNN, but also shows that visualization is not enough for deep CNN.

Introduction

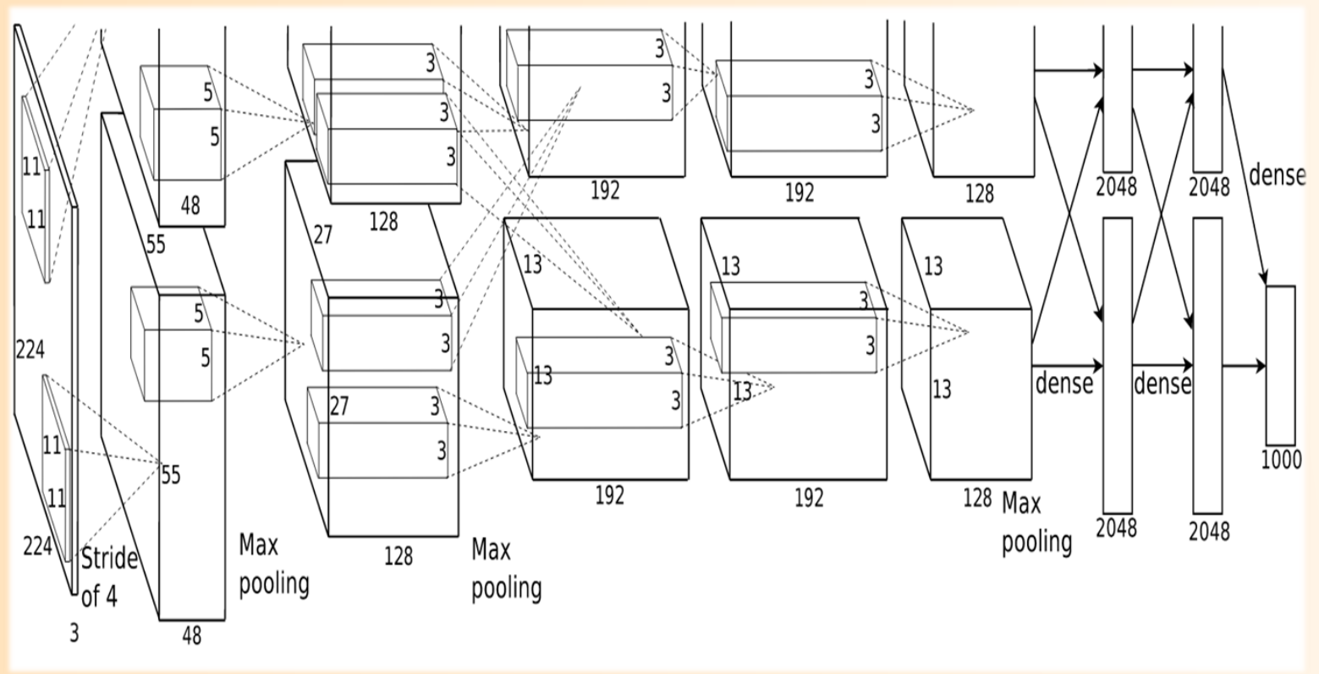
Convolutional Neural Network (CNN) was first proposed by Y. LeCun et al. in 1989, who successfully trained a CNN for digit classification. However, not until 2012, when A.Krizhevsky et al. applied this architecture in ILSVRC-2012 competition and won the first place with top-5 error rates of 15.3% (which improves by more than 10% compared to previous feature engineering methods), did CNN take over computer vision. In recent years, the advance of GPU, the availability of much larger training sets, and better model regularization strategies all contribute to the dramatic improvement in performance. Nevertheless, on the outset, it was unclear what CNN actually learned and thus cast doubt on the model. In this project, we discuss and compare different methods of visualization on various well-known models, in order to gain further sights into the structure and success of CNN.

Visualization Methods

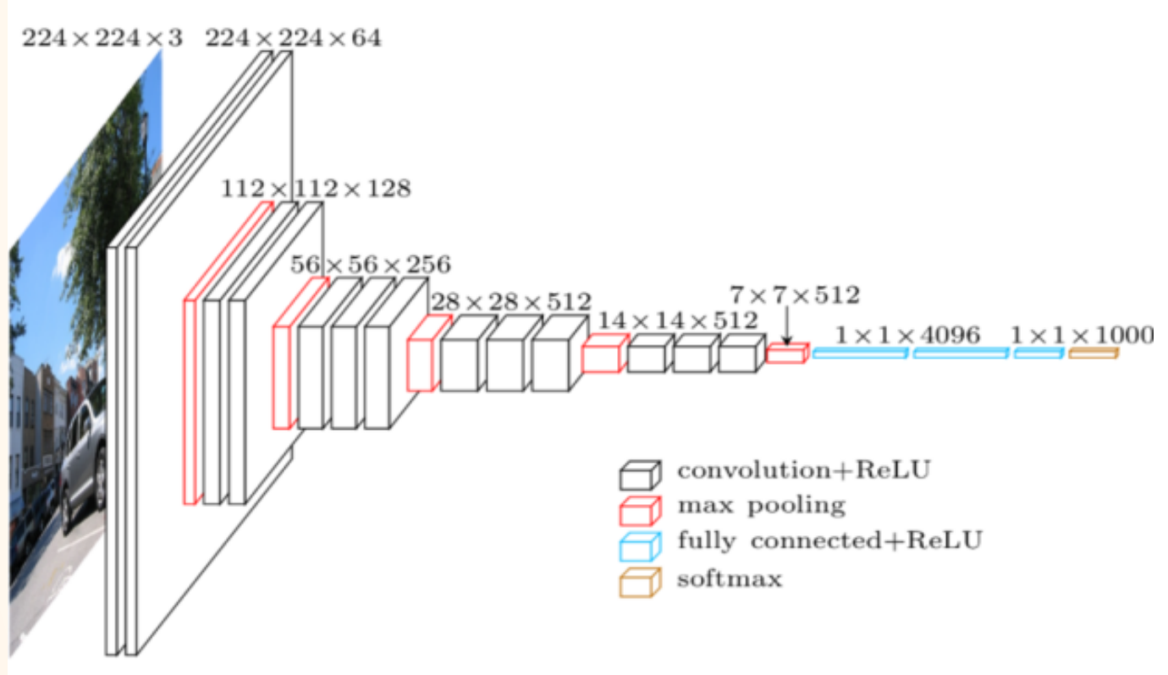
- Activity: Visualize the output of a neuron for a given image.
- Deconvolutional Network: Reconstruct the input image from a given neuron by unpooling, ReLu and deconvolution (transpose convolution).
- Saliency Map: Calculate the gradient of a score model for a class with respect to the input image.
- Deep Generator Network (DGN): Use a pretrained image generator instead of hand-crafted priors.
- Plug-and-Play Generative Networks (PPGN): Improve from DGN using denoising autoencoder to restrict the input-code space.

Models *(top-5 error rate in parentheses)*

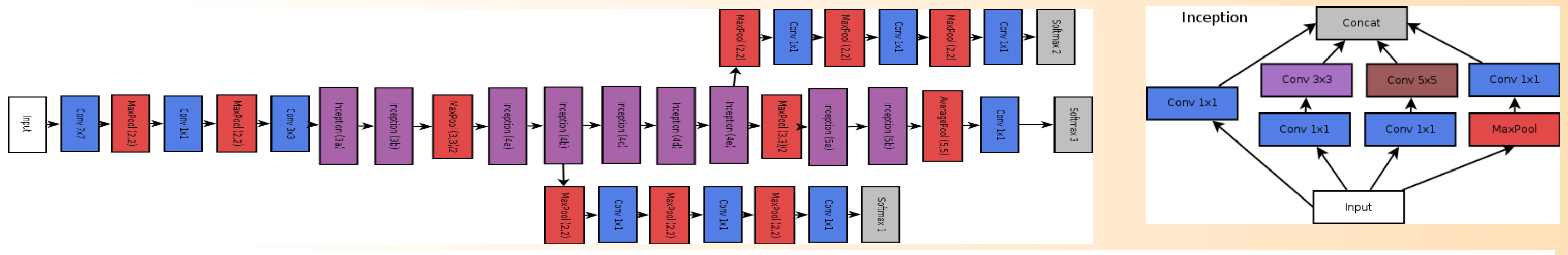
AlexNet: (15.3 %)



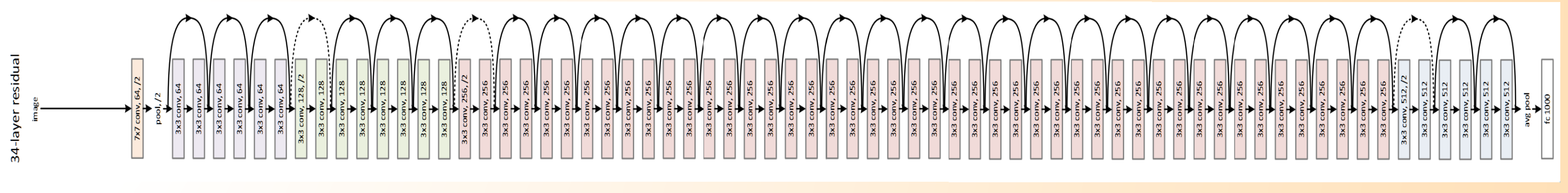
VGG: (7.3%)



GoogLeNet: (6.7%)



ResNet: (3.57%)



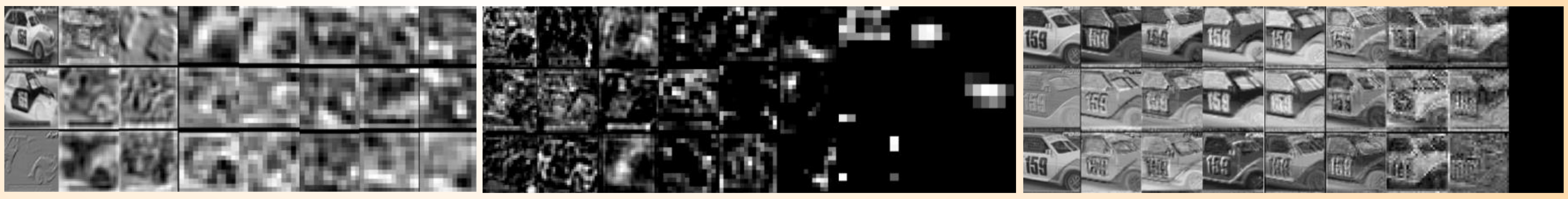
Results

Testing Images:

Wolf *Car* *Keyboard*

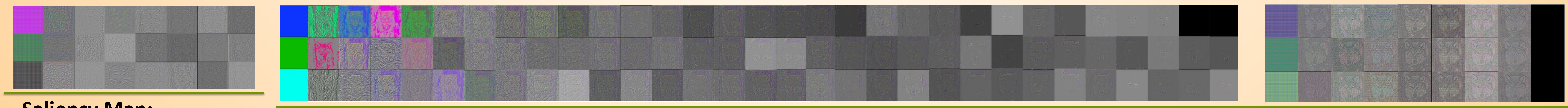
Activation:

Alexnet *GoogLeNet* *ResNet101*



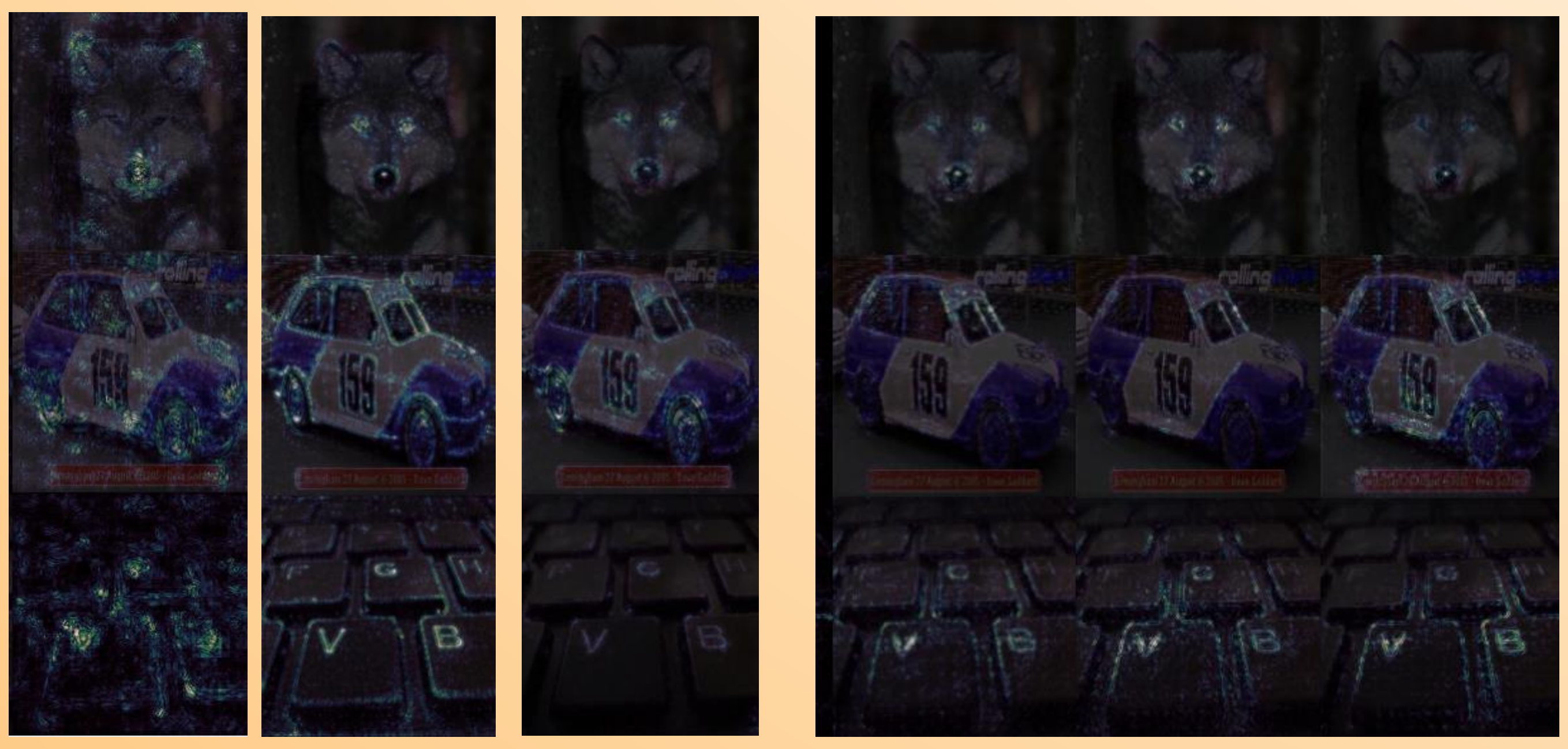
Deconvolutional Network:

Alexnet *VGG16* *ResNet50*

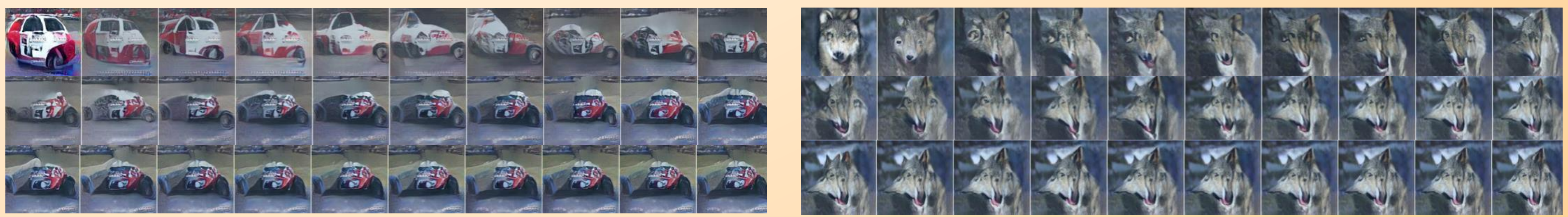


Saliency Map:

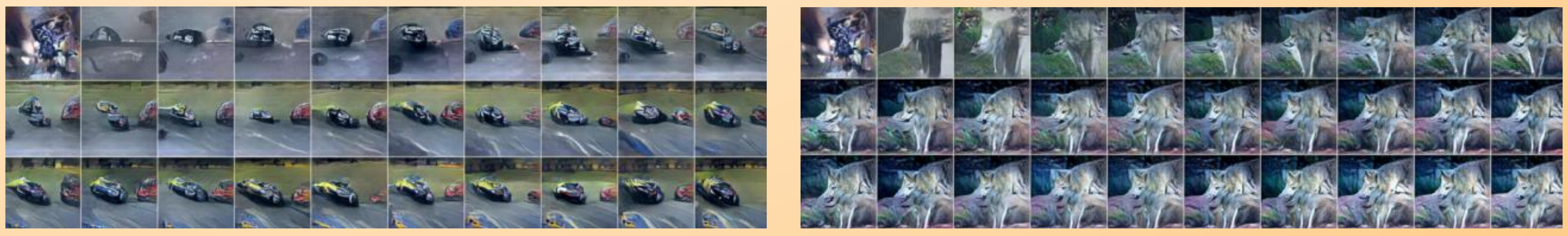
Alexnet *VGG16* *GoogLeNet* *ResNet50* *101* *152*



Deep Generator Network:



Plug-and-Play Generator Network:



Conclusion

In this project, we perform various experiments on different models to try to gain insight into what were actually learned. The results not only strengthen the fact that CNN is able to discover small patterns and integrate patterns into more and more complex structures, but also shine light into the "black box". This enables us to find shortcomings of a particular model and improve the performance. However, as CNN goes deeper and deeper, it is also obvious that visualization methods is incapable of explaining everything.

References

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel: Backpropagation applied to handwritten zip code recognition. (1989)
[2] D. Erhan, Y. Bengio, A. Courville, and P. Vincent: Visualizing higher-layer features of a deep network. (2009)
[3] M. D. Zeiler, R. Fergus: Visualizing and understanding convolutional networks. (2013)
[4] A. Krizhevsky, I. Sutskever, G. Hinton: Imagenet classification with deep convo- lutional neural networks. (2012)
[5] K. Simonyan, A. Vedaldi, A. Zisserman: Deep inside convolutional networks: Visualising image classification models and saliency maps. (2013)
[6] K. Simonyan, A. Zisserman: Very deep convolutional networks for large-scale image recognition. (2014)
[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich: Going Deeper With Convolutions. (2015)
[8] K. He, X. Zhang, S. Ren, J Sun:Deep Residual Learning for Image Recognition. (2015)
[9] A. Nguyen, J. Yosinski, J. Clune: Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. (2015)
[10] A. Nguyen, A. Dosovitskiy, J. Yosinski, T. Brox, J. Clune: Synthesizing the preferred inputs for neurons in neural networks via deep generator networks (2016)
[11] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, J. Yosinski: Plug & Play Generative Networks: Conditional Iterative Generation of Images in Latent Space. (2017)

Acknowledgement

We would like to thank Prof. Hung-yi Lee and all TAs for giving such an enriched lecture.