DeepDream: What I Dreamt While At College

Minh B. T. Nguyen*
Department of Electrical and Computer Engineering
Virginia Tech
mnguyen0226@vt.edu

1 Problem Statement

The two problems that I am trying to solve: Interpreting deep learning model prediction behaviour in computer vision, and creating eye-pleasing art with via deep learning.

Interpretability in deep learning computer vision deployment project is occasionally ignore due to the black-box implementation, the focus on accuracy-related metrics, and the lack of powerful visualization tool (1). How does one checks whether the model has correctly learned the correct features? There are several attempts to mathematically understand deep learning model's prediction such as LIME (2), Attribution (3), and Grad-CAM (4). In addition, in recent years, the usage of deep learning to create artistic imagery demanded methods such as Neural Style Transfer (5). There is a lack of tool that can both doing model interpretability while generating artistic imagery.

I am interested in Explainable AI techniques since they help me debug during deep learning architecture designing and training process. Besides, as a student, I faced several stressful situations as well as solving joyful challenges which caused me to dreamed several nights. I would like to use deep learning to visualize and share my dream.

2 Challenges

There are two main challenges: Building, training, testing ResNet model from scratch in large dataset while aiming for high accuracy; and creating eye-pleasing visualization.

3 Dataset

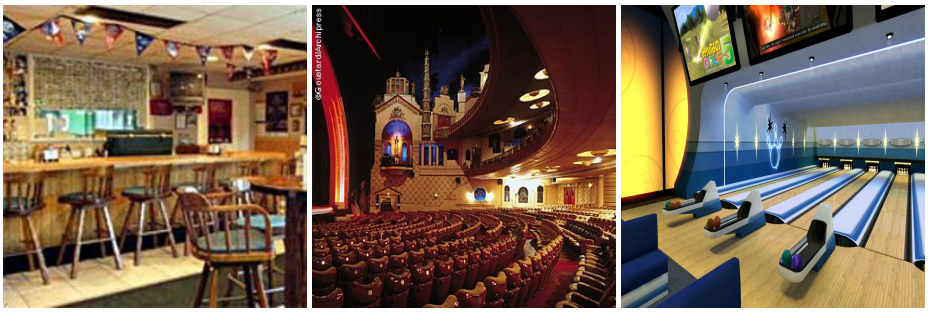

Figure 1: Samples from MIT Indoor Scenes Dataset: Bar, Auditorium, Bowling (6)

The dataset I will use is MIT Indoor Scene (7) (6). According to "MIT Indoor Scenes" dataset published on kaggle.com, the dataset contain 67 "Indoor" categories with the total of 15620 images

*Final Project Proposal.

(6). It is noted that the number of images varies across categories but there are at least 100 images per category (6). The images are RGB and in .jpg format. There are a total of (67x80) 5360 images for training and (67x20) 1340 images for testing. Since the images are in different size, I will need to do some image reshaping. I chose this dataset because it is light while complex, it is used in several research papers (8), (9), (10), and its categories match with my dream theme.

4 Method and Improvement

I will use the "DeepDream Algorithm" which is "an artistic algorithm where a pretrained CNN is fed an image and optimized to amplify the features it 'sees' in the image" (11). The main ideas is to apply "gradient ascent" at several activation function instead of gradient descent. In addition, I will use the "Image Pyramid" and "Gradient Smoothing" methods recommended by Aleksa Gordić which provide a more vivid "dreaming" results (12).

However, the current implementation (11) (12) used the pretrained models either from ImageNet (13) or Places205 (14) datasets. In this project, I will train my own ResNet architecture from scratch with MIT Indoor Scenes dataset, then I will apply the "DeepDream Algorithm". I have not decided the depth of the ResNet since it depends on the power of my local machine. The two main goals here are: providing a high-performance ResNet on MIT Indoor Scene dataset, and generating beautiful deep dream imageries.

5 Results Evaluation

The ResNet performance will be evaluated via Confusion Matrix (true positive, false positive, true negative, false negative), Train Accuracy vs Test Accuracy Per Epoch plot, Train Loss vs Test Loss Per Epoch plot, and PCA, T-SNE visualization.

In terms of artistic imageries, I expect the result will be somewhat has the styles similar to Alexander Mordvintsev's (11) and Aleksa Gordić's (12) works:

Figure 2: Samples from Gordić's and Mordvintsev's "DeepDream" results (11) (12)

References

[1] X. Li, H. Xiong, X. Li, X. Wu, X. Zhang, J. Liu, J. Bian, and D. Dou, "Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond," *arXiv preprint arXiv:2103.10689*, 2021.

[2] S. Mishra, B. L. Sturm, and S. Dixon, "Local interpretable model-agnostic explanations for music content analysis.," in *ISMIR*, vol. 53, pp. 537–543, 2017.

[3] S. Jha, S. Raj, S. Fernandes, S. K. Jha, S. Jha, B. Jalaian, G. Verma, and A. Swami, "Attribution-based confidence metric for deep neural networks," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[4] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[5] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 11, pp. 3365–3385, 2019.

[6] M. Ahmad, "Indoor scene recognition cvpr," *https://www.kaggle.com/itsahmad/indoor-scenes-cvpr-2019*, 2019.

[7] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 413–420, IEEE, 2009.

[8] A. Yang, P. M. Esperança, and F. M. Carlucci, "Nas evaluation is frustratingly hard," *arXiv preprint arXiv:1912.12522*, 2019.

[9] A. López-Cifuentes, M. Escudero-Viñolo, J. Bescós, and Á. García-Martín, "Semantic-aware scene recognition," *Pattern Recognition*, vol. 102, p. 107256, 2020.

[10] M. Rahimzadeh, S. Parvin, E. Safi, and M. R. Mohammadi, "Wise-srnet: A novel architecture for enhancing image classification by learning spatial resolution of feature maps," *arXiv preprint arXiv:2104.12294*, 2021.

[11] A. Mordvintsev, C. Olah, and M. Tyka, "Inceptionism: Going deeper into neural networks," 2015.

[12] A. Gordić, "Pytorch implementation of deepdream algorithm," *https://github.com/gordicaleksa/pytorch-deepdream*, 2019.

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[14] L. Wang, S. Guo, W. Huang, and Y. Qiao, "Places205-vggnet models for scene recognition," *arXiv preprint arXiv:1508.01667*, 2015.