CV1 Project#4

Zimo He
YuanPei Collage
2100017750@stu.pku.edu.cn

1 Deep FRAME model

The code implementation of this project can be found on https://github.com/mileret/deep_frame.

1.1 Configuration

In this task, we employed a convolutional neural network (CNN) with a three-layer structure as our deep FRAME model. Each layer of the CNN consisted of 128, 64, and 32 convolutional kernels, respectively. The final layer's feature map was summed and utilized as an activation to measure the discrepancy between the source image and the generated image.

1.2 Qualitative Results

The results generated by the model with aforementioned configuration are presented in Fig. 1. We compared the generated results of CNN models with 1, 2, and 3 layers, respectively. From the visualization results, we can infer that an increase in the number of model layers corresponds to an improvement in the quality of generated outputs. Besides, we also visualize the kernels of the CNN model's first layer across different images, which is shown in Fig. 2.

1.3 Ablation Study

Furthermore, we conducted an ablation study that substantiates the necessity of the model's size. Specifically, we utilized a smaller CNN with three layers, consisting of 16, 8, and 4 convolutional kernels, respectively. The generated results of this smaller model are displayed in Fig. 3. It is evident from the results that the utilization of a smaller model leads to a decline in the quality of the generated outputs.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

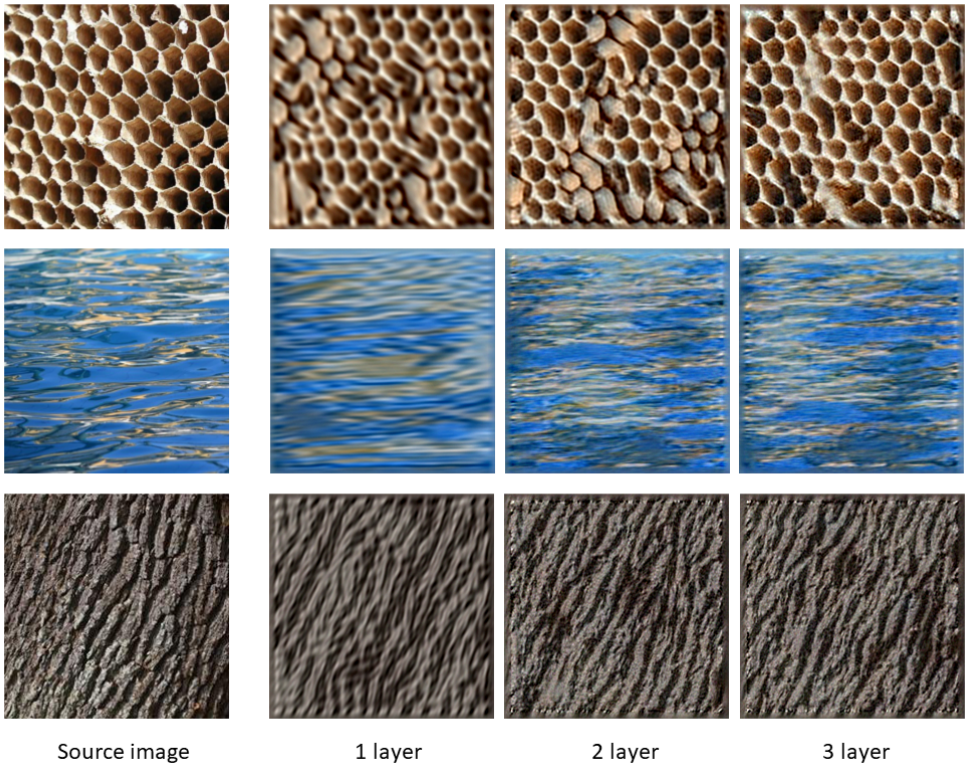

Figure 1: This figure illustrates the results generated by the deep FRAME model based on the source image. From left to right, the images represent the source image, the results generated by the 1-layer model, the 2-layer model, and the 3-layer model, respectively. It is evident from these images that an increase in the number of model layers corresponds to an improvement in the quality of generated outputs.

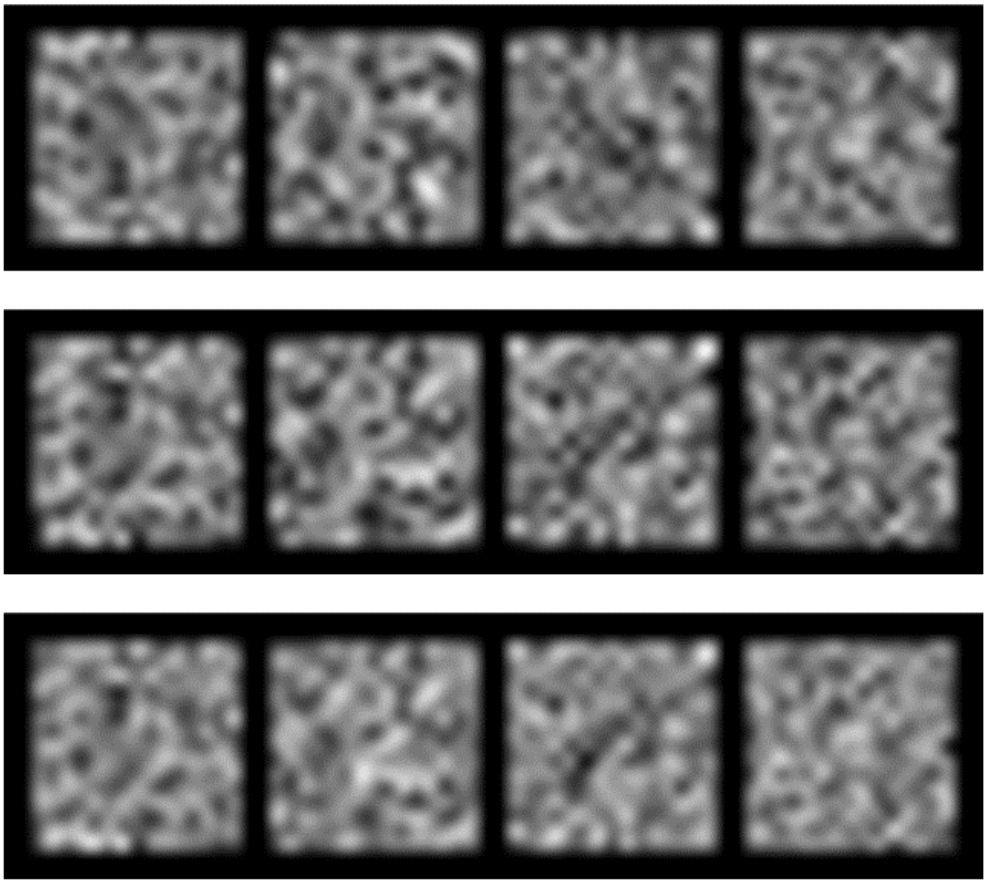

Figure 2: The figure presented showcases the visualization of the first four kernels of the CNN model's first layer. From top to bottom, the kernels are learned from images of beehive, water, and bark, respectively. Through visual examination of the results, it seems that the first layer's kernels have learned texture-based features of the images.

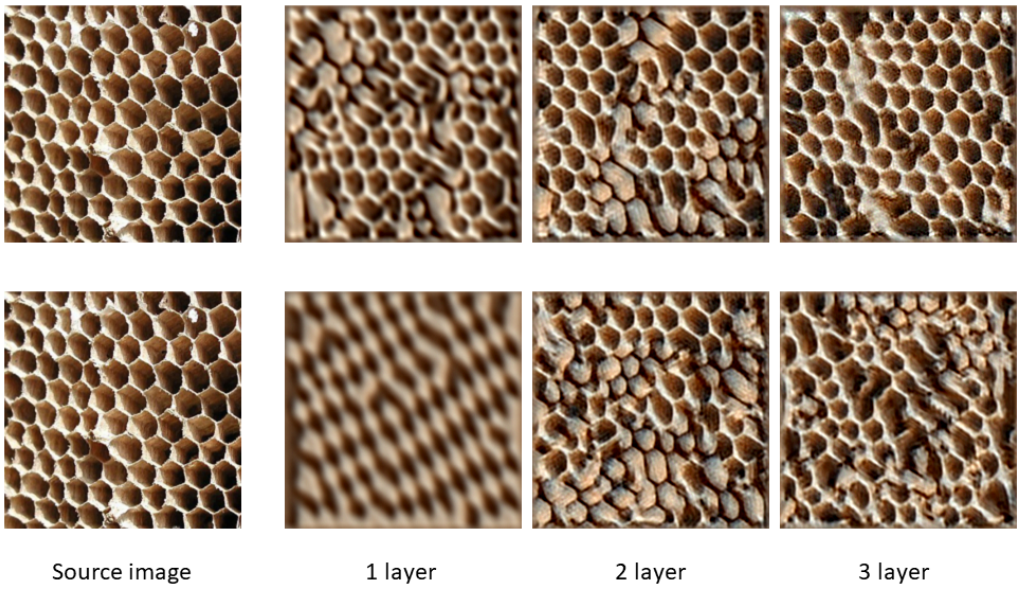

Figure 3: The figure presented illustrates the results generated using a smaller CNN model. The first row displays the results generated by the model with the structure described in the Configuration section. The second row showcases the results generated by the smaller CNN model.