

Deep learning applications

Deep Learning Practical Work

Aymeric DELEFOSSE & Charles VIN

2023 – 2024



Contents

1	Transfer Learning	3
1.1	VGG16 Architecture	3
1.2	Transfer Learning with VGG16 on 15 Scene	3
1.2.1	Approach	3
1.2.2	Feature Extraction with VGG16	4

normaliser les images par m sig de image net
3.

Chapter 1

Transfer Learning

1.1 VGG16 Architecture

1. ★ **Knowing that the fully-connected layers account for the majority of the parameters in a model, give an estimate on the number of parameters of VGG16.** There are three fully-connected layers at the end of the VGG16 architecture. Calculating their weights is relatively straightforward. We take into account the inclusion of biases.

- The first fully-connected layer receives an input of size 7 by 7 by 512 (the resulting output of the convolutional layers), which equals an input size of 25,088. Knowing that there are 4,096 neurons, this layer has a total of $(25,088 + 1) \times 4,096 = 102,764,544$ trainable weights.
- The second fully-connected layer receives the input from the previous layer, which is 4,096, and also consists of 4,096 neurons, resulting in $(4,096 + 1) \times 4,096 = 16,781,312$ trainable weights.
- Lastly, the third fully-connected layer, consisting of 1,000 neurons, has $(4,096 + 1) \times 1,000 = 4,097,000$ parameters.

Thus, the fully connected layers account for a total of 123,642,856 parameters. We can confidently state that they represent at least 85% of the model, implying there should be around 140 million parameters to learn. If we consider a margin of 5%, there should be between 137,380,951 and 154,553,570 parameters.

We can readily confirm that the convolutional layers account for 14,714,688 parameters, meaning that there are actually 138,357,544 parameters in VGG16, meaning the fully-connected layers accounts to 89% of parameters.

2. ★ **What is the output size of the last layer of VGG16? What does it correspond to?** The output size of the last layer of VGG16 is 1000. It corresponds to the 1000 classes of the ImageNet dataset that the model has been trained on. Each element in this output vector represents the network's prediction scores for a specific class in the ImageNet dataset, and the class with the highest score is considered the predicted class for our given input image.

3. **Bonus: Apply the network on several images of your choice and comment on the results.** bóbr

4. **Bonus: Visualize several activation maps obtained after the first convolutional layer. How can we interpret them?**

1.2 Transfer Learning with VGG16 on 15 Scene

1.2.1 Approach

5. ★ **Why not directly train VGG16 on 15 Scene?** The 15 Scene dataset is quite small compared to the massive ImageNet dataset that VGG16 was originally trained on. VGG16 requires a lot of data to generalize well and to avoid overfitting. Moreover, training such a model from scratch is computationally expensive and time-consuming.

6. ★ **How can pre-training on ImageNet help classification for 15 Scene?** What's the assumption behind transfert learning : the data are somehow simililar

7. What limits can you see with feature extraction?

1.2.2 Feature Extraction with VGG16

8. What is the impact of the layer at which the features are extracted?

9. The images from 15 Scene are black and white, but VGG16 requires RGB images. How can we get around this problem?

10. Rather than training an independent classifier, is it possible to just use the neural network? Explain.