**Google Net Architecture**

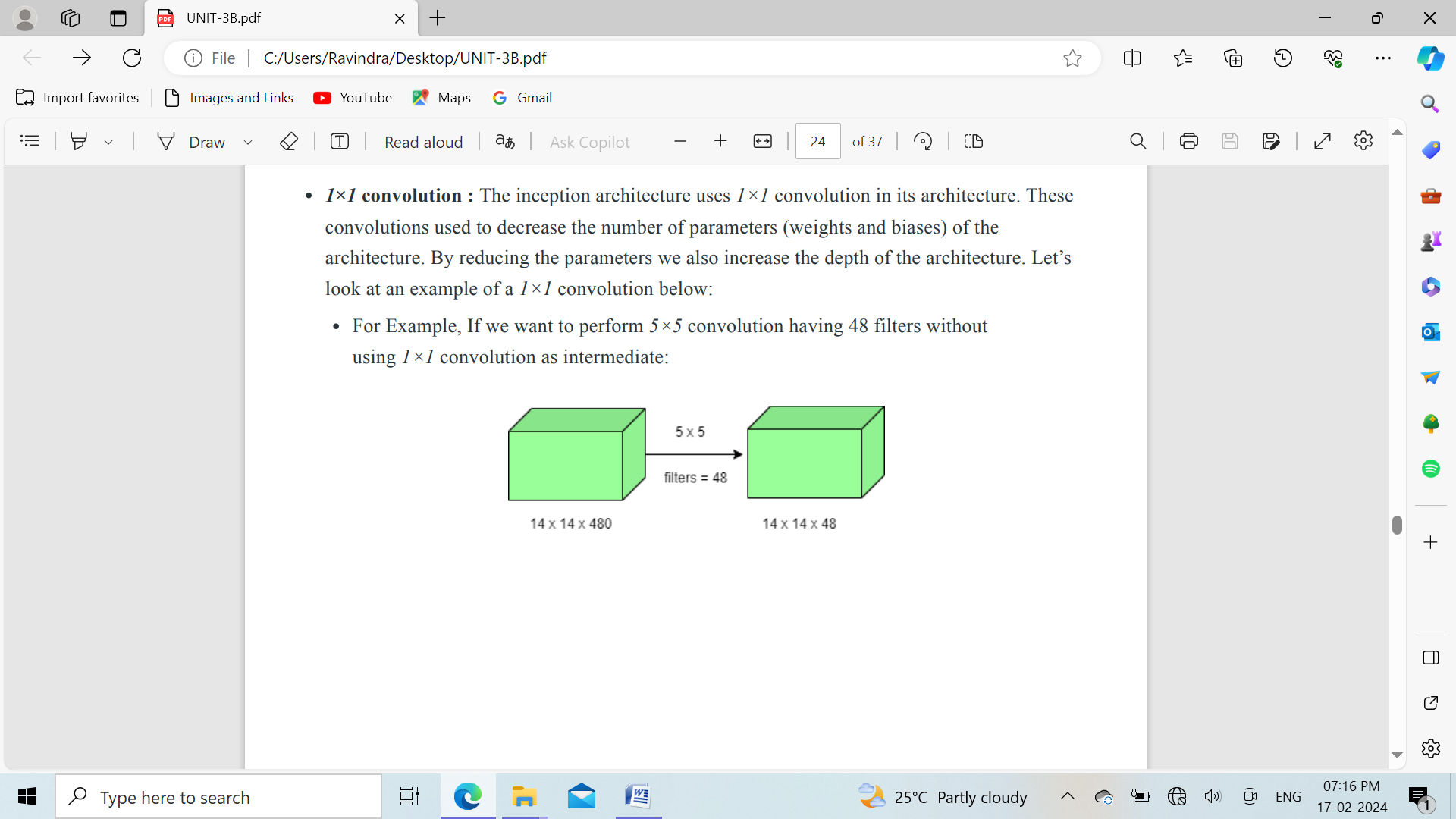
GoogleNet is the first version of Inception Models (or Inception V1) was proposed by research at Google (with the collaboration of various universities) in 2014 in the research paper titled “Going Deeper with Convolutions”. This architecture was the winner at the ILSVRC 2014 image classification challenge. It has provided a significant decrease in error rate as compared to previous winners AlexNet (Winner of ILSVRC 2012) and significantly less error rate than VGG (2014 runner up).

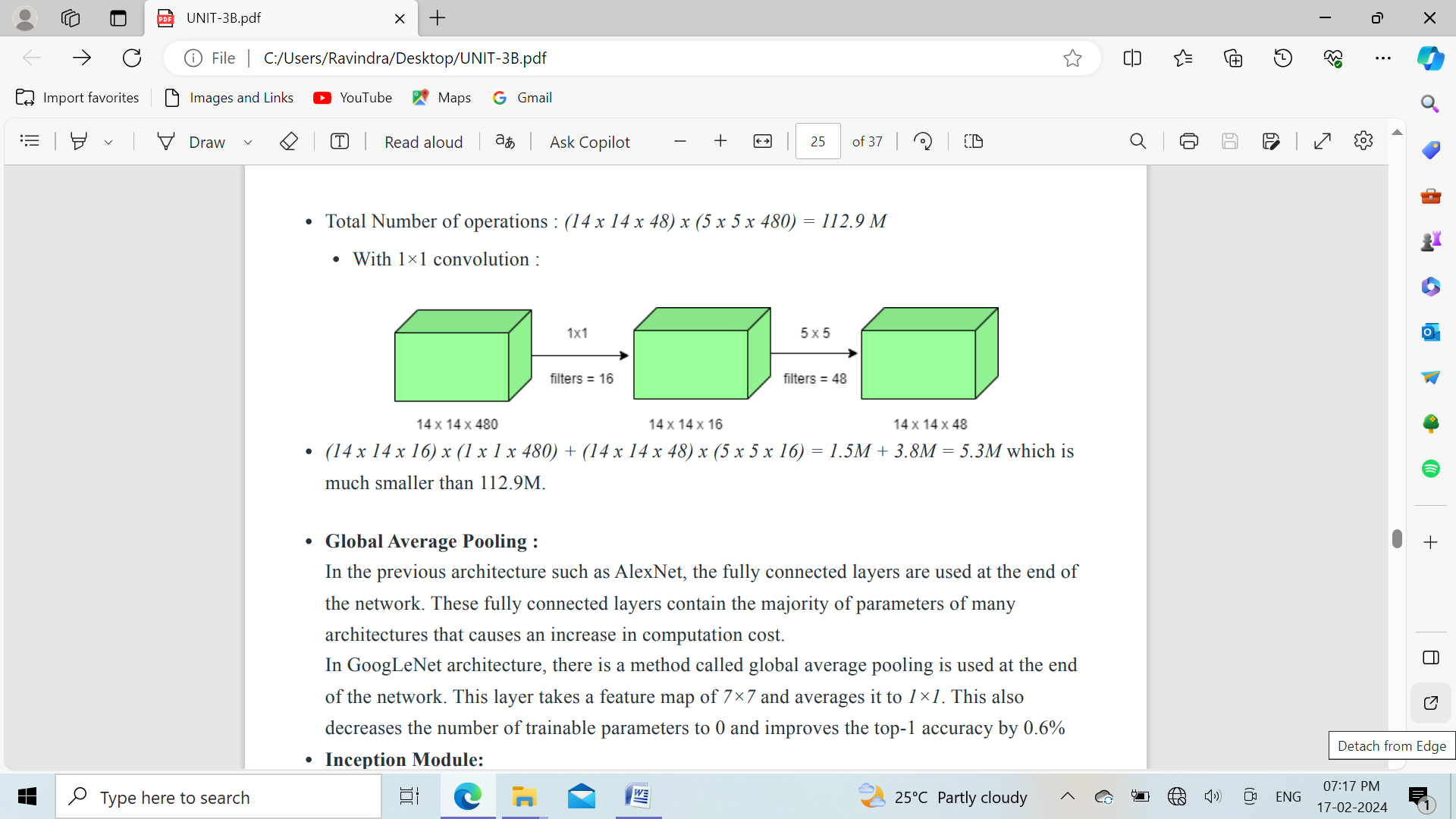
**Features of GoogleNet:**

This architecture uses techniques such as 1×1 convolutions in the middle of the architecture and global average pooling.

**1×1 convolution :** The inception architecture uses 1×1 convolution in its architecture. These convolutions used to decrease the number of parameters (weights and biases) of the architecture. By reducing the parameters we also increase the depth of the architecture.

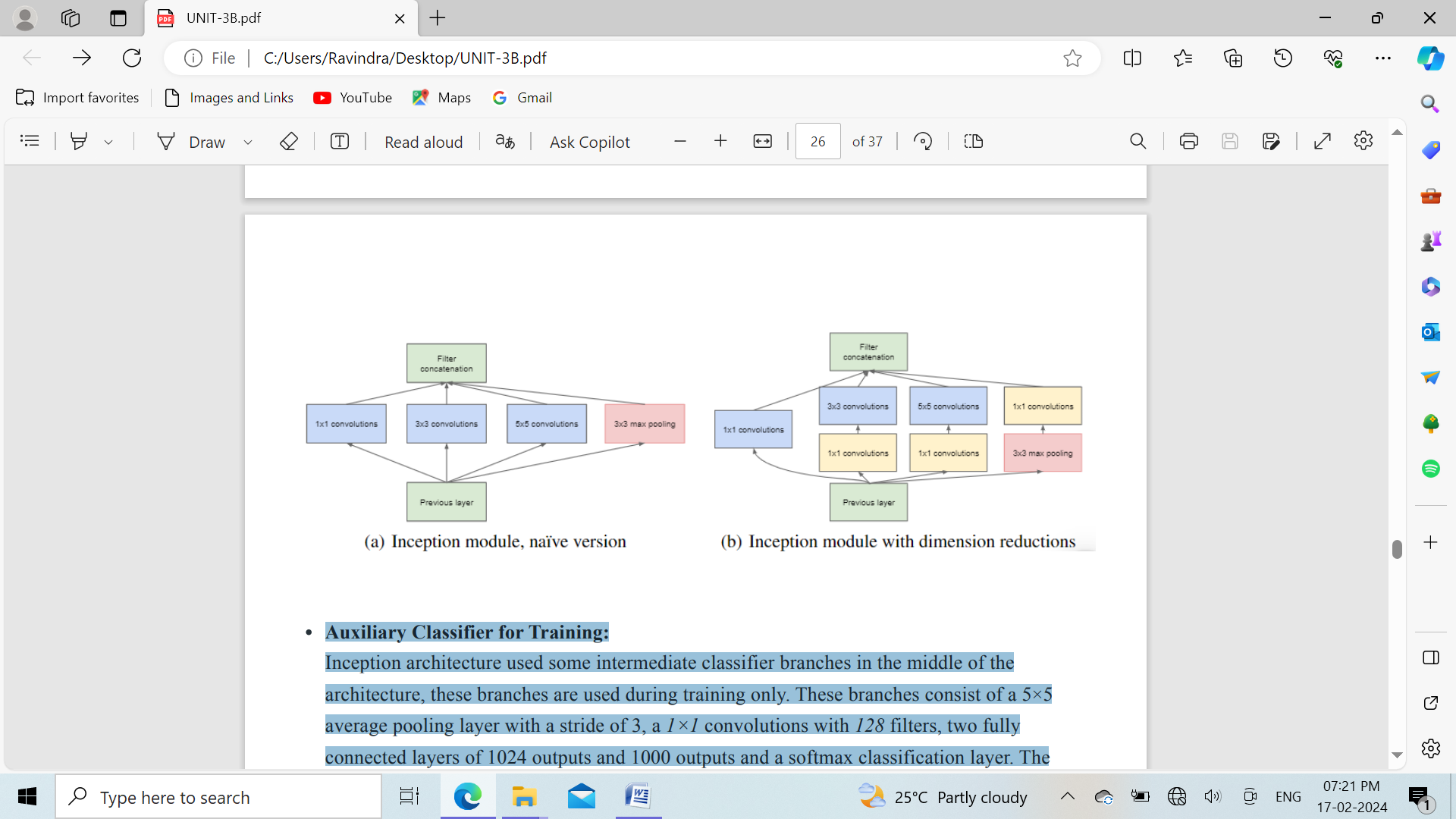
Let’s look at an example of a 1×1 convolution below: For Example, If we want to perform 5×5 convolution having 48 filters without using 1×1 convolution as intermediate:

****

****

**Global Average Pooling :** In GoogLeNet architecture, there is a method called global average pooling is used at the end of the network. This layer takes a feature map of 7×7 and averages it to 1×1. This also decreases the number of trainable parameters to 0 and improves the top-1 accuracy by 0.6%

**Inception Module:** The inception module is different from previous architectures such as AlexNet, ZF-Net. In this architecture, there is a fixed convolution size for each layer. In the Inception module 1×1, 3×3, 5×5 convolution and 3×3 max pooling performed in a parallel way at the input and the output of these are stacked together to generated final output. The idea behind those convolution filters of different sizes will handle objects at multiple scales better.



**Auxiliary Classifier for Training:** Inception architecture used some intermediate classifier branches in the middle of the architecture, these branches are used during training only. These branches consist of a 5×5 average pooling layer with a stride of 3, 1×1 convolutions with 128 filters, two fully connected layers of 1024 outputs and 1000 outputs and a softmax classification layer. The generated loss of these layers added to total loss with a weight of 0.3. These layers help in combating gradient vanishing problem and also provide regularization.

**Model Architecture:**

The overall architecture is 22 layers deep. The architecture was designed to keep computational efficiency in mind. The idea behind that the architecture can be run on individual devices even with low computational resources. The architecture also contains two auxiliary classifier layer connected to the output of Inception (4a) and Inception (4d) layers.

The architectural details of auxiliary classifiers as follows:

* An average pooling layer of filter size 5×5 and stride 3.
* A 1×1 convolution with 128 filters for dimension reduction and ReLU activation.
* A fully connected layer with 1025 outputs and ReLU activation
* Dropout Regularization with dropout ratio = 0.7
* A softmax classifier with 1000 classes output similar to the main softmax classifier.

This architecture takes image of size 224 x 224 with RGB color channels. All the convolutions inside this architecture uses Rectified Linear Units (ReLU) as their activation functions.

**Results:** GoogLeNet was the winner at ILSRVRC 2014 taking 1 place in both classification an detection task. It has top-5 error rate of 6.67% in classification task. An ensemble of 6 GoogLeNets gives 43.9 % mAP on ImageNet test set.