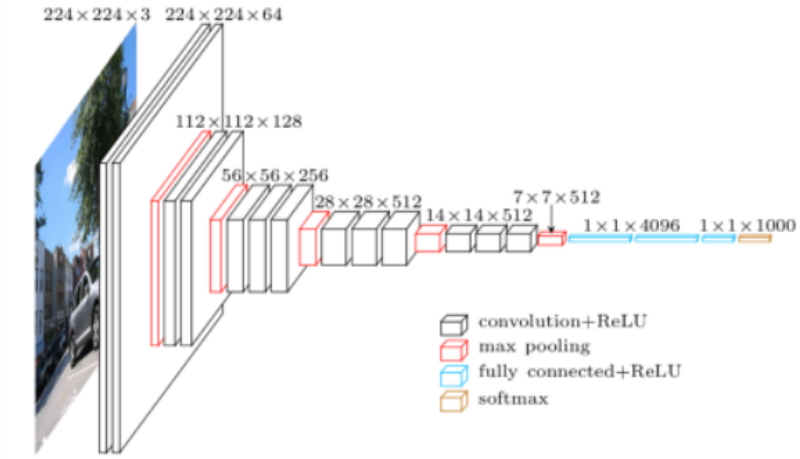
Visual Geometry Group Implementation Issues

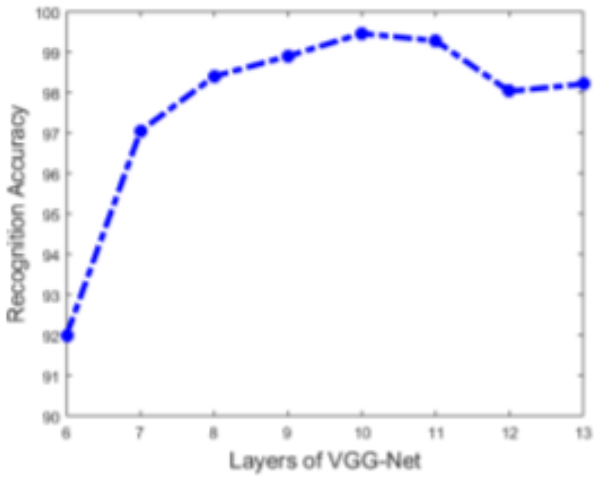
Visual Geometry Group (VGG) is an image classification convolutional neural network developed by a team, of the same name, at the University of Oxford. This network is specially characterized by its pyramidal shape, the bottom layers (closer to the image) are wide and the top layers are deep. As seen in the figure below, VGG consists of subsequent convolutional layers followed by pooling layers. The pooling layers are responsible for making the layers narrower.



Given an image, the VGG network will output probabilities of the different classes that an image could possibly belong to. For example, Ria inputted an image to the network. Then, let us say the VGG network outputs a value of 0.94 for dog and 0.03 for cat. This indicates that there is a 94% chance, we can say with 94% confidence, that the image contains a dog and there is a 3% chance, we can say with 3% confidence, that the image contains a cat.

The following are the implementation issues associated with VGG network:

- Image-quality covariates affects the performance of VGG. Gaussian blur and JPEG compression has a significant effect on the performance of this model as it causes a quick drop in the performance with an increase in standard deviation of the Gaussian filters (Grm, Struc, Artiges, Caron, Ekenel, 2017).
- Gaussian and salt-and-pepper noise also negatively impact the performance of the VGG network. Noise is an important factor affecting this model and sufficiently low levels of noise need to be assured for reliable verification performance.
- The VGG network is also affected by missing data. Missing information especially around the periocular region, followed by the eye, nose, and mouth regions have a negative impact on this model's performance.

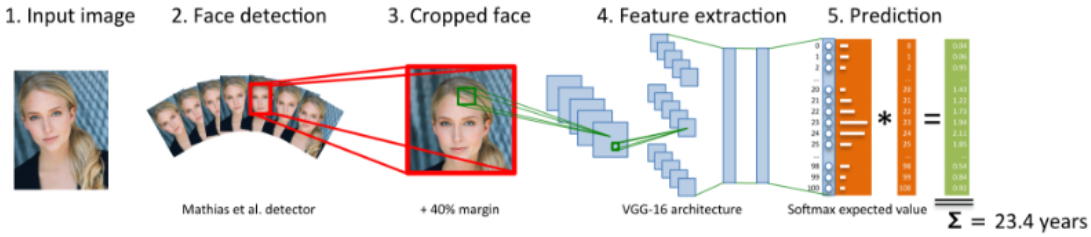


The figure above shows the recognition accuracy for features of different layer of VGG network for a sample dataset

- Brightness distortion also has an impact on the VGG network's

performance. An increase in brightness results in a poor performance of this model. Among all the deep learning models for face recognition, VGG performs the worst when the brightness is increased.

- This model has many weight parameters, the models are very heavy with almost 550 MB + of weight size. This results in a long inference time.
- This model has 19.6 billion FLOPs which makes it complex. This value is much higher than that for other models like AlexNet and ResNet.
- This model has hidden fc (fully-connected) layers (more than ResNet) which gives us a good accuracy with training data but performs poorly with testing data.
- The VGG-Face network's last few layers are fully connected. Since our dataset is small, fine-tuning the pretrained network on a small dataset might lead to overfitting.



The figure above shows us how VGG is implemented to estimate the apparent age in still face images

- A convolution significantly slows down an operation. Since the VGG network is deep, each training step is going to take much longer.

- In the presence of class imbalanced data, for our model, training is very irregular. It goes down for a few batches and abnormally high for a few batches, without ever going back to normal.
- VGG is computationally expensive since it has a large number of parameters due to the presence of the last few dense layers.

Grm, K. Struc, V. Artiges, A. Caron, M. Ekenel, H. Strengths and Weaknesses of Deep Learning Models for Face Recognition Against Image Degradations. 2017. Retrieved from: https://arxiv.org