# UNDERSTANDING CLOUDS FORMATIONS

## Using Deep Learning and CNNs

Empirical Research Project
Andrea Corvi, AA 2019-2020 UCSC

# THE PROJECT

- Objective: To classify clouds photos taken by NASA Worldiew satellite.
- Problem to solve:  Multilabel classification.
- Available Data: More than 5000 images for a total of 11 thousands cloud formations.

# Shallow Clouds

The clouds that have to be classified are of a particular kind called "shallow" clouds.

Their are part of the family of the stratocumulus clouds, created by weak convective currents.

But why are they important?

# Shallow Clouds

The clouds that have to be classified are of a particular kind called "shallow" clouds".

Their are part of the family of the stratocumulus clouds, created by weak convective currents.

But why are they important?

Advanced Review

## Cloud feedback mechanisms and their representation in global climate models

Paulo Ceppi,[1*] Florent Brient,[2] Mark D. Zelinka[3] and Dennis L. Hartmann[4]

*Edited by Eduardo Zorita, Domain Editor and Editor-in-Chief*

# Shallow Clouds

The clouds that have to be classified are of a particular kind called "shallow" clouds".

Their are part of the family of the stratocumulus clouds, created by weak convective currents.

But why are they important?

Advanced Review

**Cloud feedback mechanisms and their representation in global climate models**

Paulo Ceppi,[1*] Florent Brient

*Edited by Eduardo Zorita, Domain Edi*

GEOPHYSICAL RESEARCH LETTERS, VOL. 31, L24212, doi:10.1029/2004GL021188, 2004

**Climatic shift in patterns of shallow clouds over the Amazon**

F. J. F. Chagnon, R. L. Bras, and J. Wang

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Received 29 July 2004; revised 21 October 2004; accepted 19 November 2004; published 24 December 2004.

# Shallow Clouds

The clouds that have to be classified are of a particular kind called "shallow" clouds".

Their are part of the family of the stratocumulus clouds, created by weak convective currents.

But why are they important?

Advanced Review

Cloud feedback mechanisms and their representation in global climate models

Paulo Ceppi ,[1*] Florent Brient

Edited by Eduardo Zorita, Domain Edi

GEOPHYSICAL RESEARCH LETTERS, VOL. 31, L24212, doi:10.1029/2004GL021188, 2004

Climatic shift in patterns of shallow clouds over the Amazon

F. J. F. Chagnon, R. L. Bras, and J. Wa
Department of Civil and Environmental Engineering,
Massachusetts, USA

Received 29 July 2004; revised 21 October 2004; acc

Environmental Research Letters

LETTER

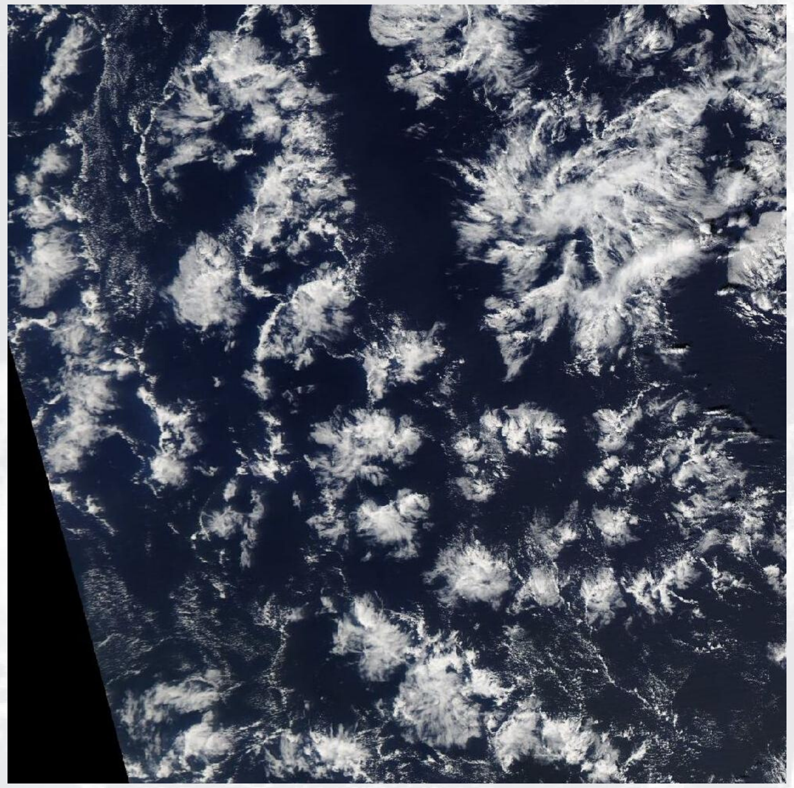
Feedback mechanisms of shallow convective clouds in a warmer climate as demonstrated by changes in buoyancy

Basically shallow clouds play an important role for the global energy balance for two main reasons:

- They reflect a significant portion of the incoming solar radiation, while contributing only marginally to the greenhouse effect. Thereby they cool our planet.
- Shallow cumulus cover large fractions of our planet's subtropical oceans (Bony et al., 2004).

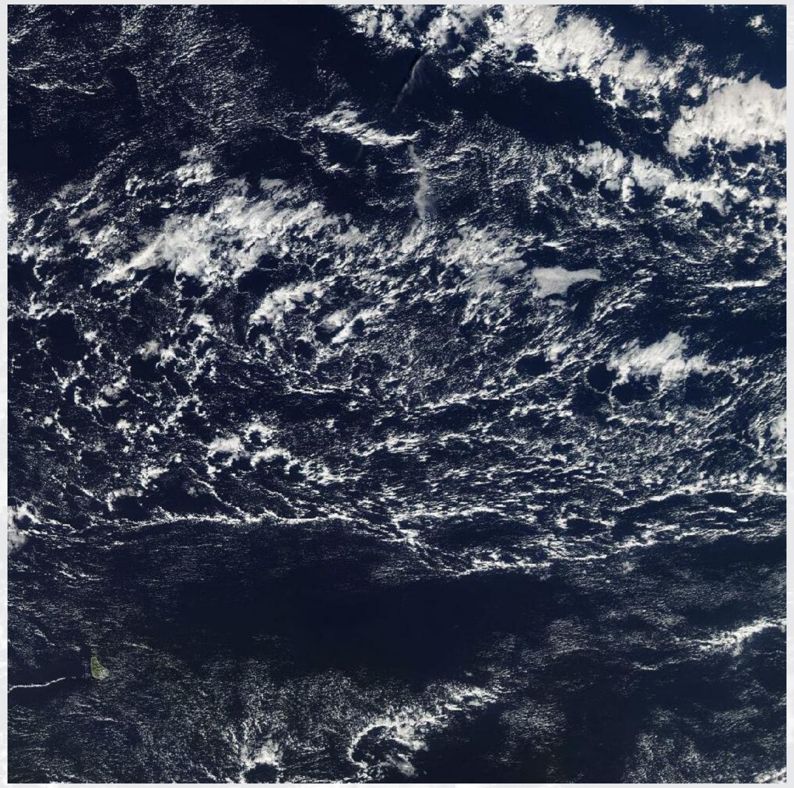
Recognising these clouds formations is then particularly important both to track their evolution due to the climate change and  to achieve a better global forecast modelling.

# The Clouds



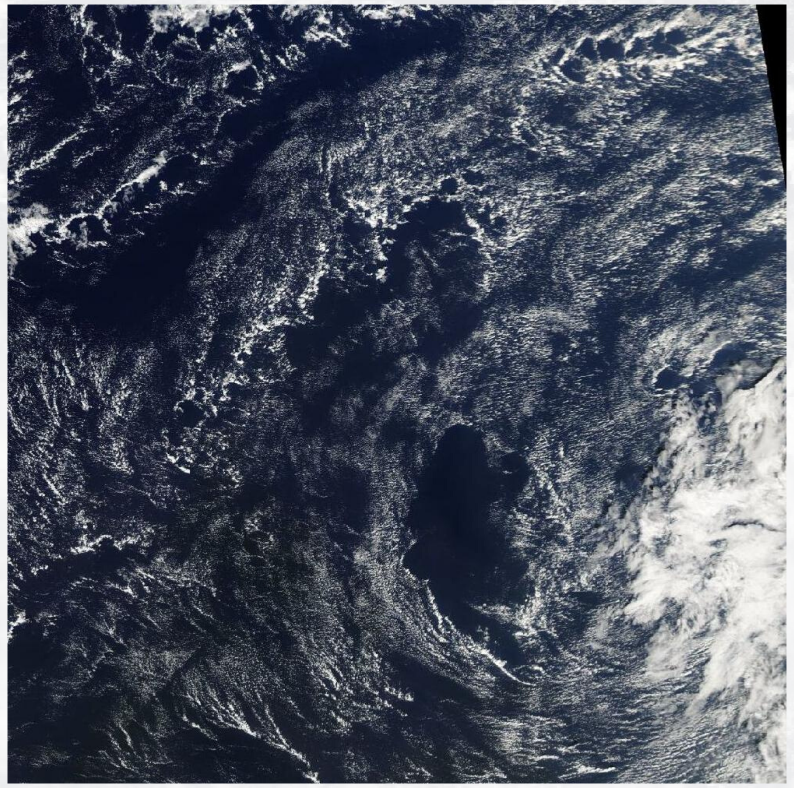
## Flower

Large-scale stratform cloud features appearing in boquets, well separated from each other.



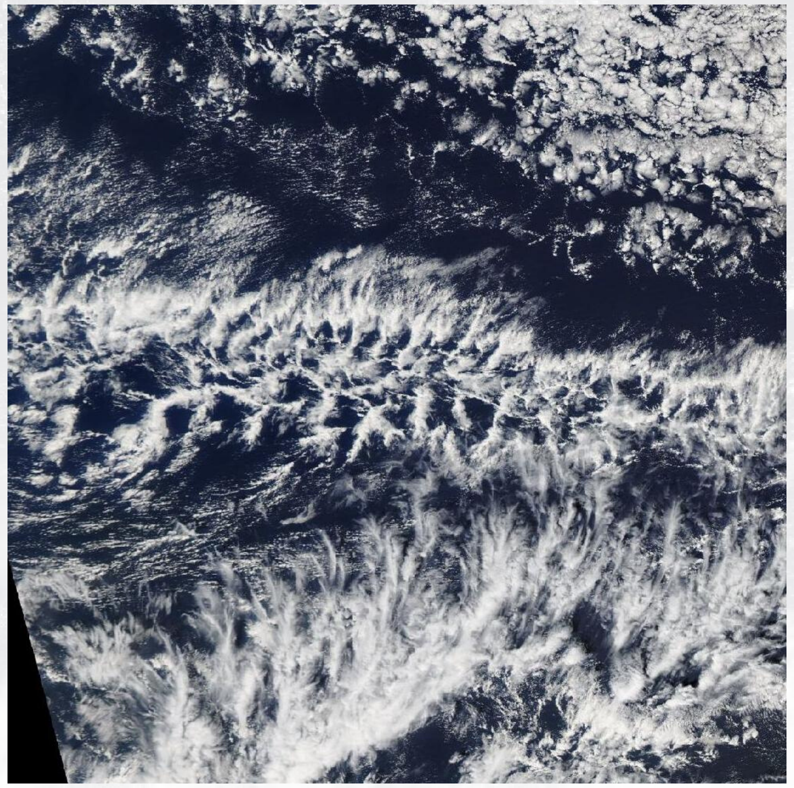
## Gravel

Meso-beta lines or arcs defining randomly interacting cells with intermediate granularity.



## Sugar

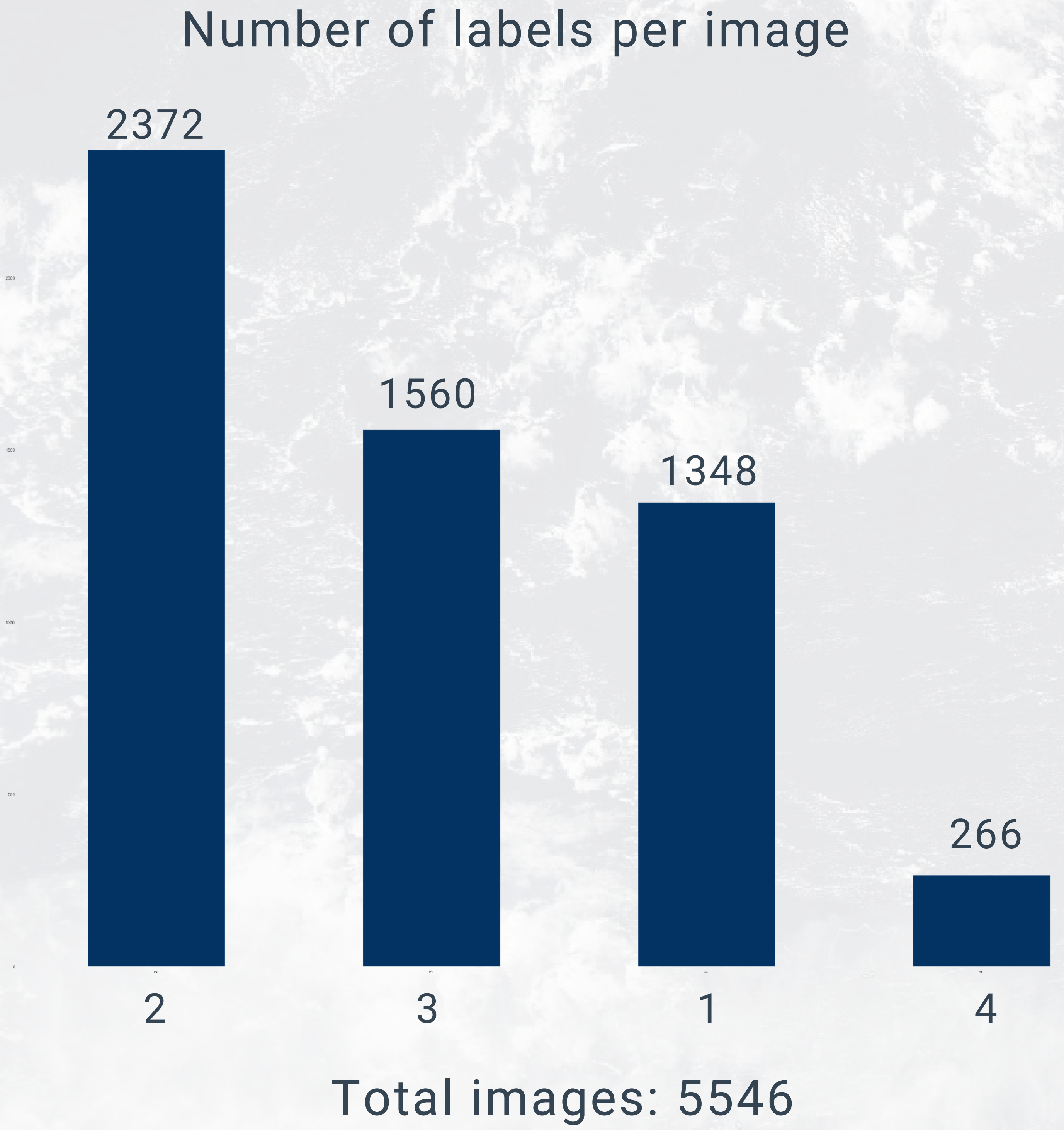
Dusting of very fine clouds, little evidence of self-organization.

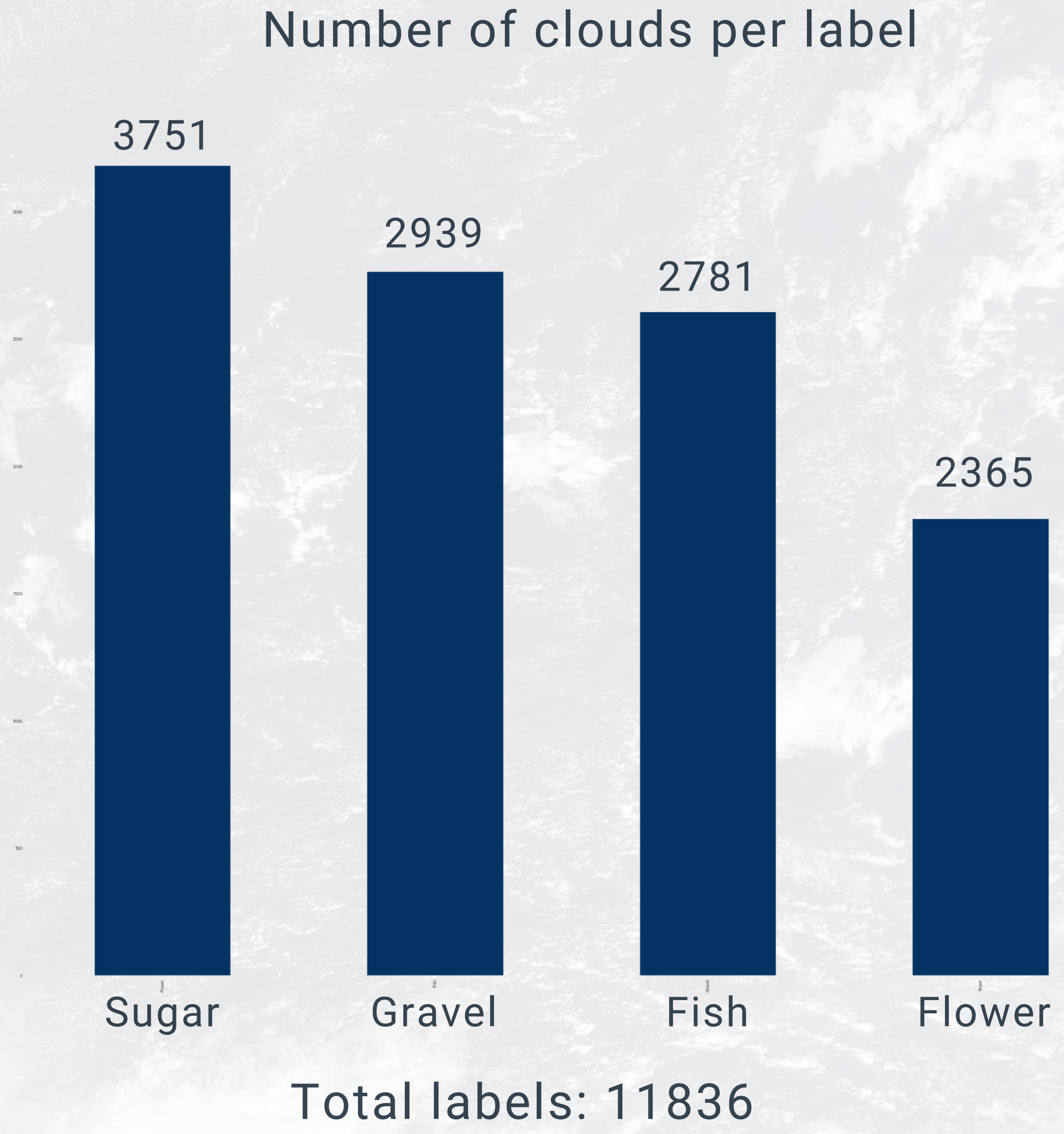


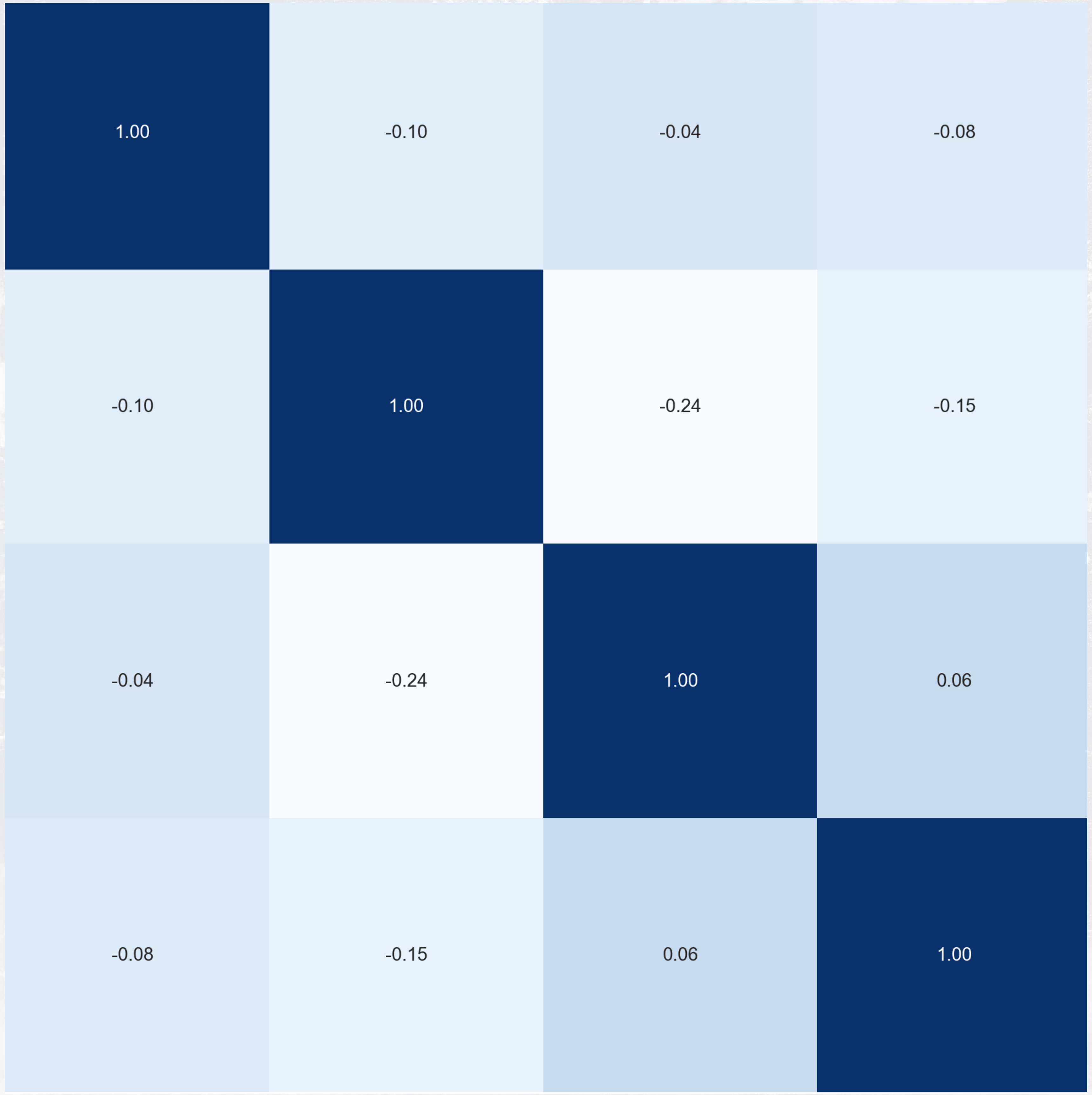
## Fish

Large-scale skeletal networks of clouds separated from other cloud forms.

# Let's have a look at the data






# Correlation Matrix

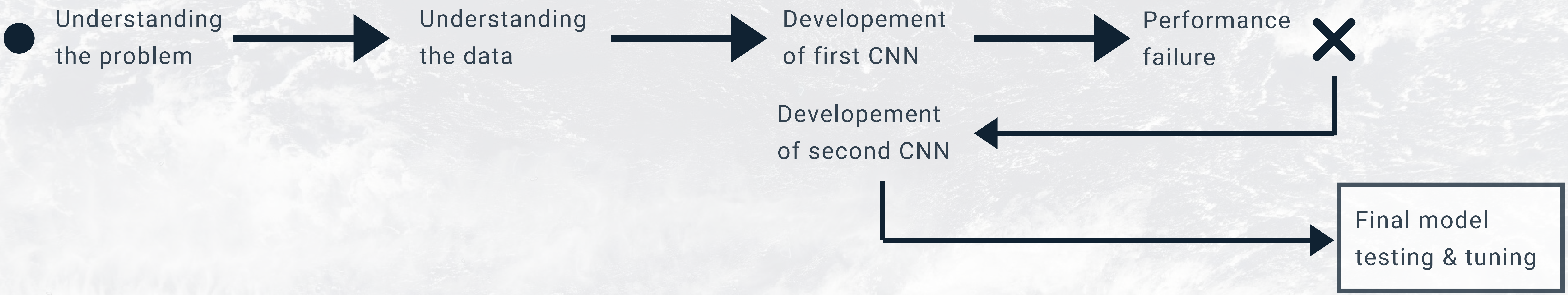
It appears that there is not any correlation between the different kind of clouds, i.e. it is not possible to say that a certain type of cloud is more likely to appear alongside another.

Back to the problem: how to tackle a multilabel classification with images as inputs?

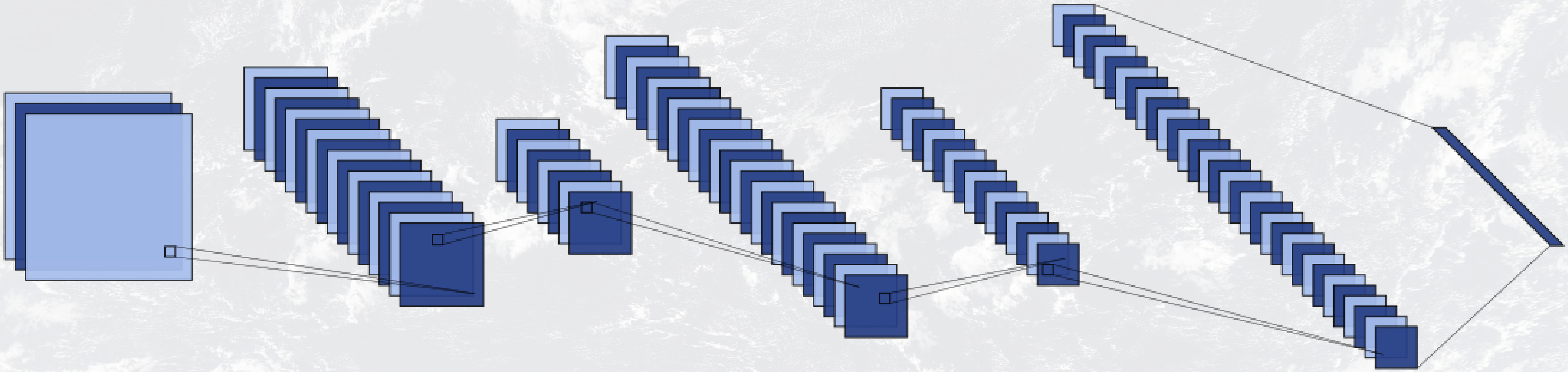
The methodology I chose for the project is the deployment of a Convolutional Neural Network.

To achieve the best results possible, firstly I implemented a "normal" CNN, but the results weren't promising since the network wasn't able to capture enough features of the clouds.

In order to obtain a better performance I decided to implement a more complex NN, already built in keras and with the weights pre-trained on ImageNet, a huge database with more than 14 millions labelled images.



# First Convolutional Neural Network



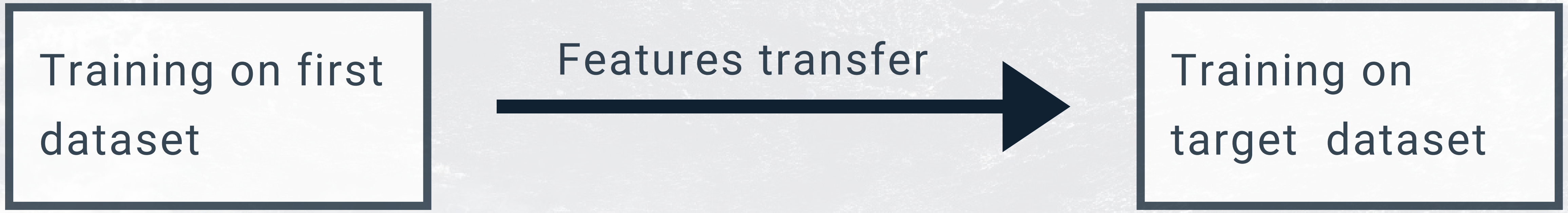
The first solution I tried was a basic convolutional neural network with a series of convolution layers then followed by pooling layers. For graphical purposes not all the layers of the net has been displayed in the image.

The problem with the net was the model was not able to capture the features in a reasonable amount of epoch, given the computational power at my disposal.

What to do then?

# Transfer Learning

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task.
Firstly a base network is trained on a base dataset and task, and then the learned features are transfered to a second target network to be trained on a target dataset and task.



# Second Convolutional Neural Network - Inception-ResNet v2

To exploit the advantages of transfer learning I decided to implent one of the NN ready to be deployed in the Keras library. The NN is called: Inception-ResNet v2, with weights trained on ImageNet.
Now, how's the architecture of such network made?

First of all it's a sort of "hybrid" CNN since it takes features from both "Inception Networks" and "Residual Networks". Both are developed by Google in order to solve different problems:

*Inception Neural Network*

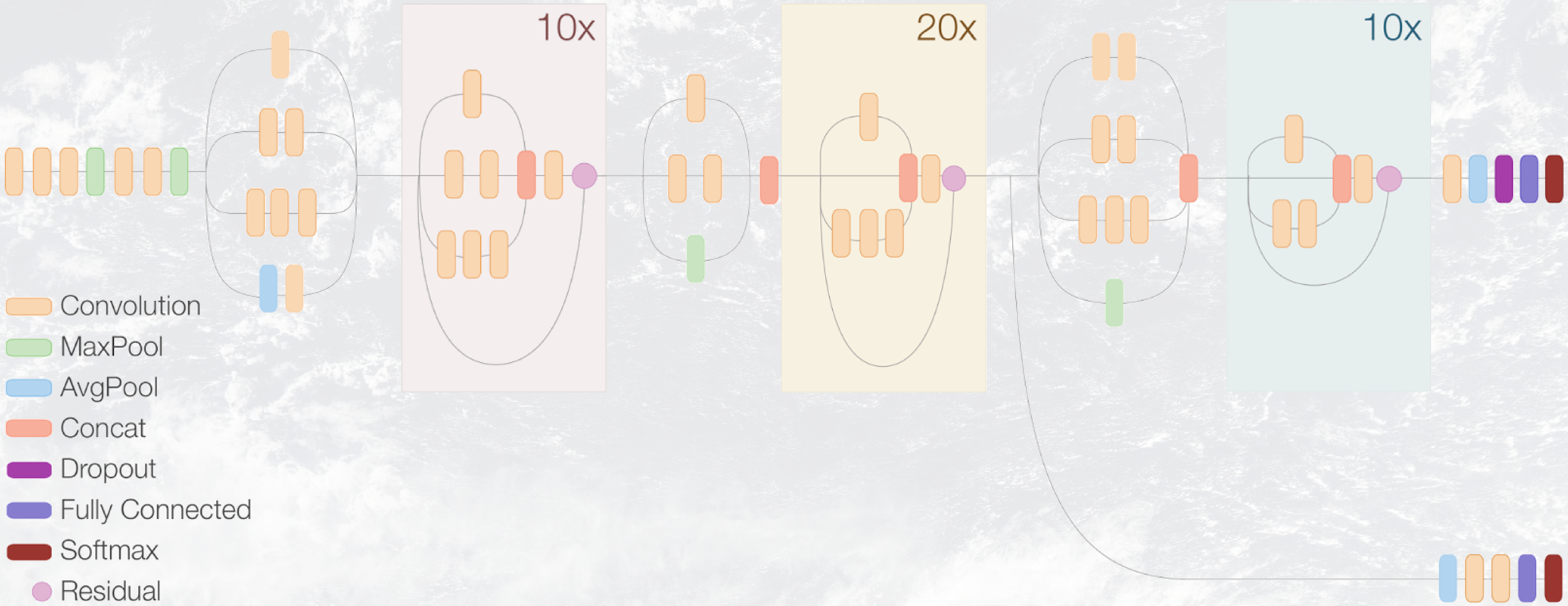
- Important parts in the image can have extremely large variation in size.
- Choosing the right kernel size becomes difficult. A larger kernel is preferred for information distributed more globally, while a smaller kernel is preferred for information that is distributed more locally.
- Very deep networks are prone to overfitting, furthermore it is computationally expensive.

*Residual Neural Network*

- They solve the vanishing gradient problem. This is because when the network is too deep, the gradients from where the loss function is calculated easily shrink to zero after several applications of the chain rule. This leads to the weights never updating their values and therefore, no learning is being performed.

# Inception-ResNet v2

How does this particular network achieve those results?




Google AI Blog

# Inception-ResNet v2

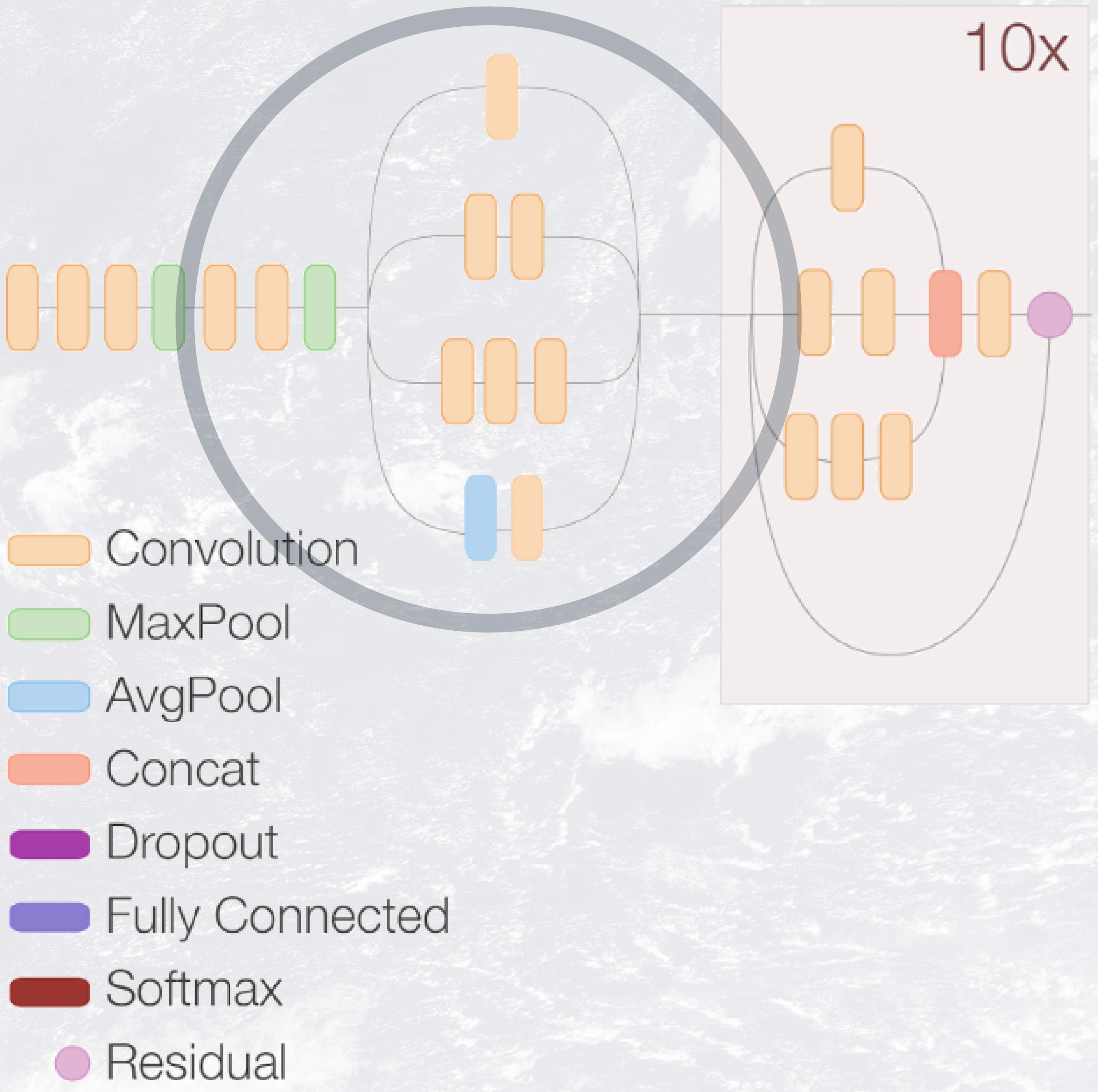
To understand it, we can focus on this part.

The outlined section is called an "Inception module" and basically exploits the following idea:

Why not have filters with multiple sizes operate on the same level? The network essentially would get a bit “wider” rather than “deeper”.

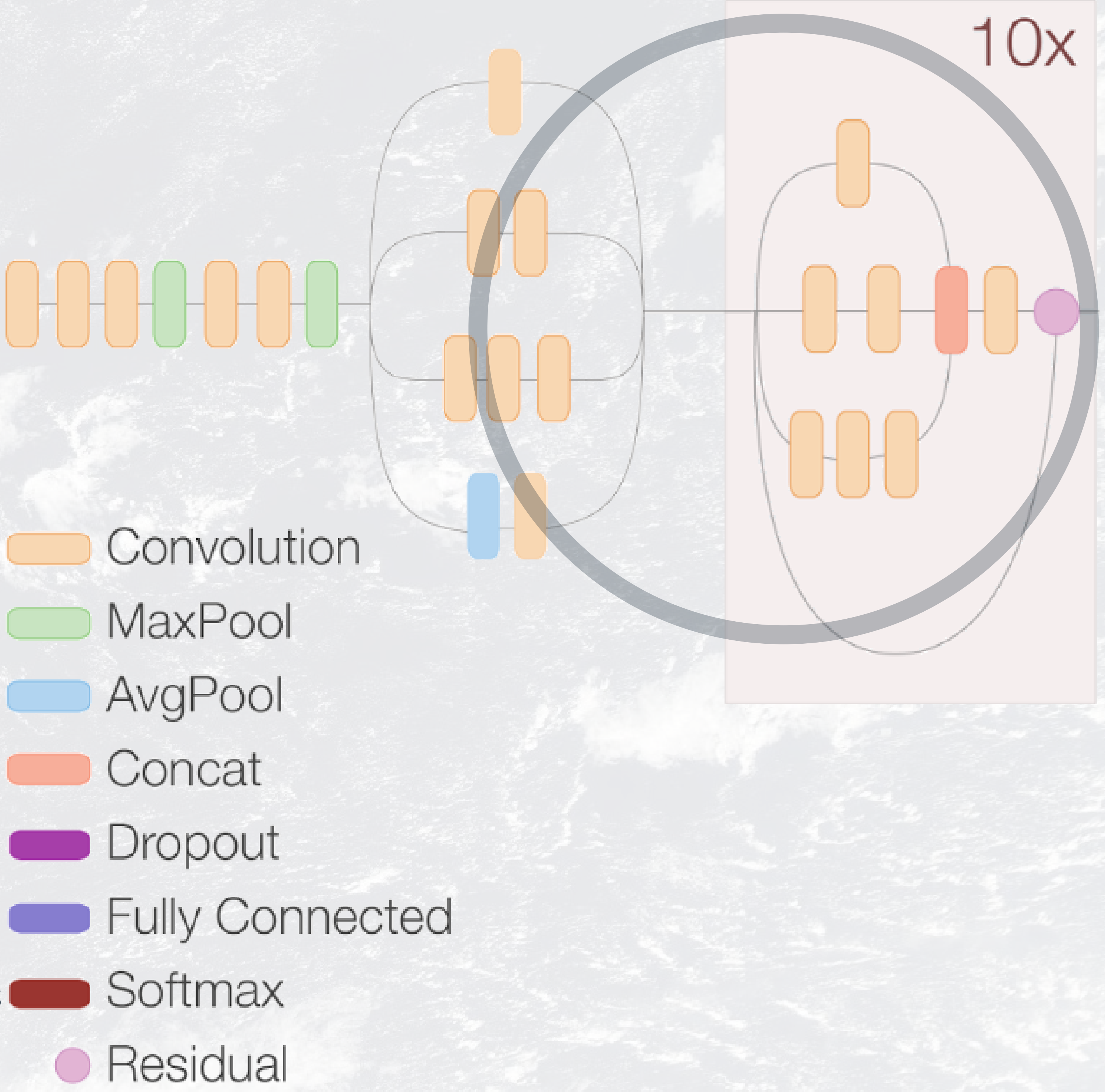
The *Inception module* does this by performing different convolution on the same input and then stacking the features into what is called a filter concatenation layer.

This clearly regards the "Inception" side of the network, but what about the "Residual" side?



# Inception-ResNet v2

To understand it, we can focus on this part.
The outlined section is called an "Inception module" and basically exploits the following idea:
Why not have filters with multiple sizes operate on the same level? The network essentially would get a bit “wider” rather than “deeper”.
The *Inception module* does this by performing different convolution on the same input and then stacking the features into what is called a filter concatenation layer.
This clearly regards the "Inception" side of the network, but what about the "Residual" side?




Now *Residual Connections* are introduced (pink circle); they add the input to the output of the convolution operation of the inception module. Basically it preserves information across layers, allowing the training of very deep neural networks.

# Implementation

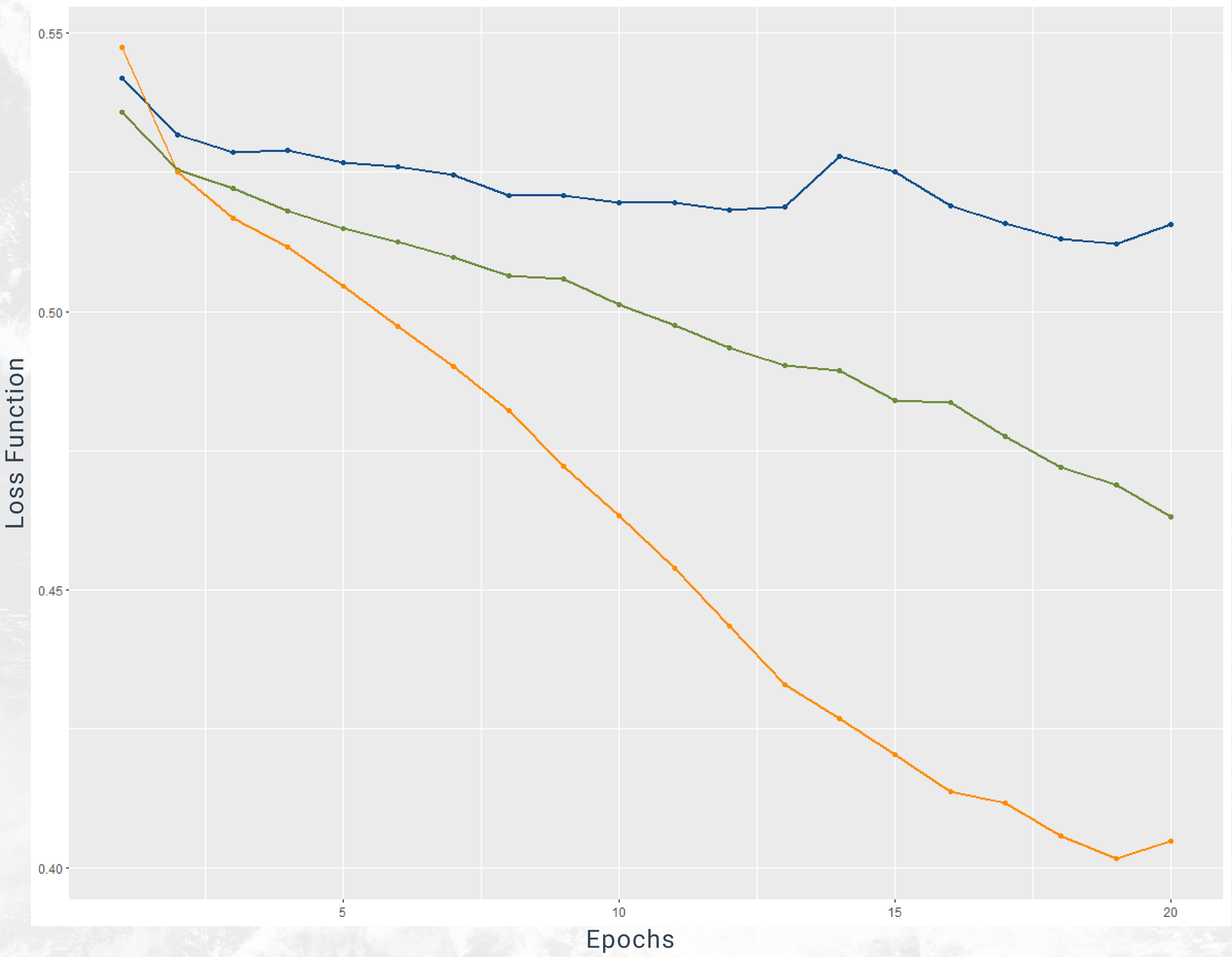
| | | |
|---|---|---|
| Activation last layer | → | Sigmoid |
| Loss function | → | Binary Crossentropy |
| Evaluation metric | → | Accuracy |
| Learning rate | → | Choosen through gridsearch |
| Number of epochs | → | Choosen through gridsearch |

# Choosing the learning rate

Learning rate:

- 0.001
- 0.0005
- 0.0001

From the graphic it seems that the smaller learning rate could be better since it reduces the loss, without stalling and has the higher accuracy. But, is it overfitting?

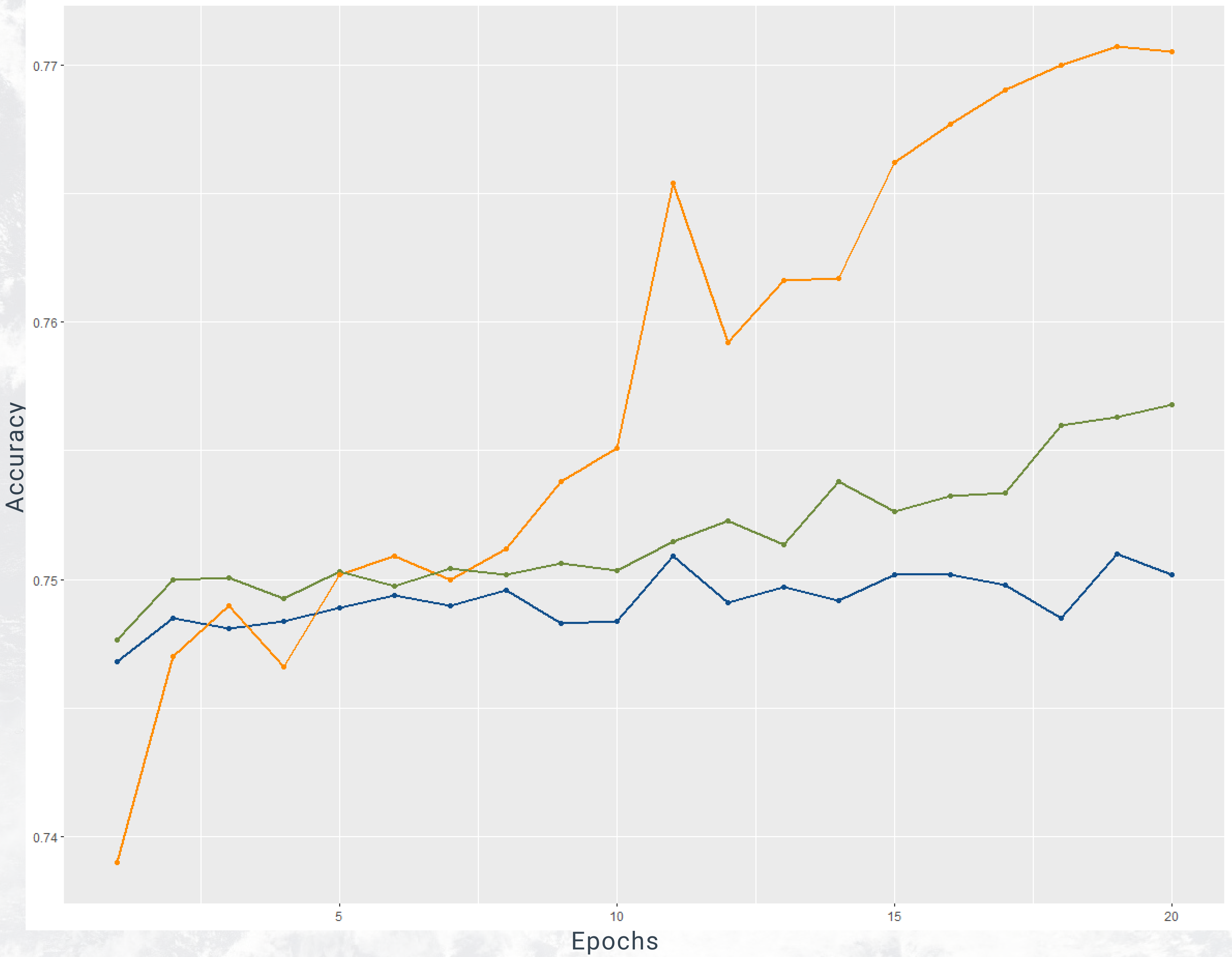


# Choosing the learning rate

Learning rate:

- 0.001
- 0.0005
- 0.0001

From the graphic it seems that the smaller learning rate could be better since it reduces the loss, without stalling and has the higher accuracy. But, is it overfitting?



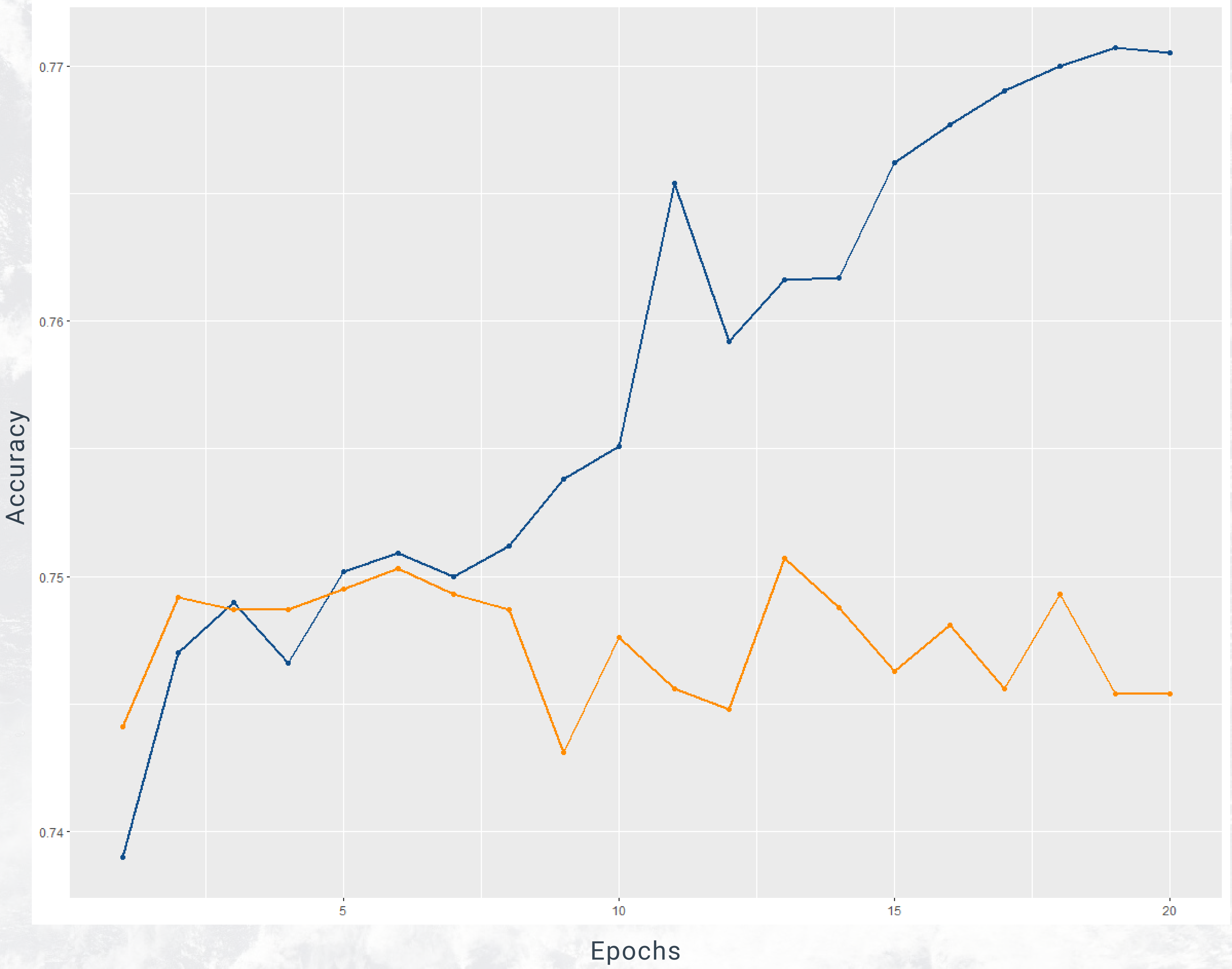
# Train Accuracy vs Test Accuracy, LR= 0.0001

Accuracy:

- Train (80%)
- Test (20%)

As expected the model will overfit at a certain epoch, so the best choice would be to stop around the fifth epoch.

Notice that although the model overfits, the span between train and test accuracy is very small.



## Limits

- Although accuracy isn't bad, it still isn't great.
- Model improvement on test data stalls after not many epoch.

## Future Developements

- To try a bigger gridsearch for the learning rate.
- Understand if it is possible to achieve similar results with simpler models.
- Add image segmentation

## Sources


- Danesi Ivan Luciano, Perugini Federica, Data Exploration for Data Science, EDUCatt, Milano, 2018.
- Google AI Blog: https://ai.googleblog.com/2016/08/improving-inception-and-image.html
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition, arXiv, 2015.
- Keras documentation.
- Simonyan Karen, Andrew Zisserman,Very deep convolutional networks for large-scale image recognition, arXiv, 2015.
- Szegedy Christian, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, arXiv, 2016.
- Kaggle.com: https://www.kaggle.com/c/understanding_cloud_organization/overview (Data source)
- Kaggle.com: https://www.kaggle.com/prazhant/an-exploratory-journey-into-the-clouds
  https://www.kaggle.com/mobassir/keras-efficientnetb2-for-classifying-cloud
- Rasp Stephan, Hauke Schulz, Sandrine Bony, Bjorn Stevens, Combining crowd-sourcing and deep learning to understand meso-scale organization of shallow convection, arXiv, 2019.
- TowardsDataScience.com:

https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8
https://towardsdatascience.com/multi-label-classification-and-class-activation-map-on-fashion-mnist-1454f09f5925