# Deep Residual Learning for Image Recognition

Leonardo Hügens

Faculdade de Ciências, Universidade do Porto

November 10, 2023

# Paper

This presentation is based on the paper Deep Residual Learning for Image Recognition by He, Zhang, Ren & Sun, from Microsoft Research.

**Deep Residual Learning for Image Recognition**

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun
Microsoft Research
{kahe, v-xiangz, v-shren, jiansun}@microsoft.com

# Outline

# Introduction

- Deeper neural networks are more difficult to train.

- The paper presents a *residual* learning framework to ease the training of networks that are substantially deeper than those used previously.

# Introduction

- Deeper neural networks are more difficult to train.

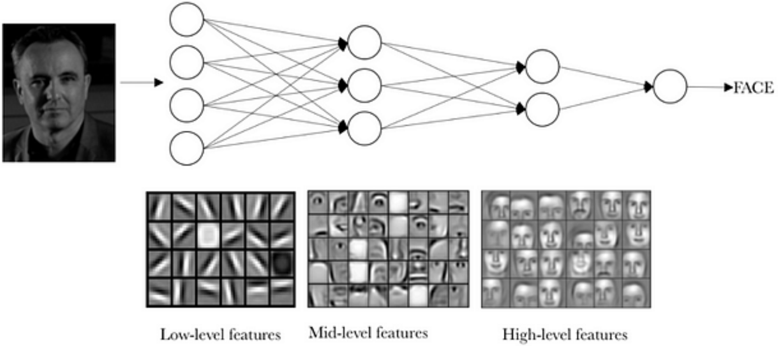
- The paper presents a *residual* learning framework to ease the training of networks that are substantially deeper than those used previously.

- The layers are reformulated as learning residual functions with reference to the layers inputs, instead of learning unreferenced functions.

- These networks are:
    - Easier to optimize
    - Can gain accuracy from considerably increased depth.

# Introduction

In a deep neural network (DNN) the "levels" of features can be enriched by the number of stacked layers.



# Introduction

The leading results on the challenging ImageNet dataset all exploit "very deep" models, with depth up to 30.

# Problem

**Question**: Is learning better networks as easy as stacking more layers?

# Problem

**Question**: Is learning better networks as easy as stacking more layers?

**Problem**: Vanishing/Exploding gradients.

**Solution**: Normalized initialization (e.g. for `ReLU` activation functions, the weights of the layer are `He` initialized, drawn from a Gaussian distribution with mean $\mu = 0$ and variance $\sigma = \frac{2}{\#inputs}$), and intermediate normalization layers.

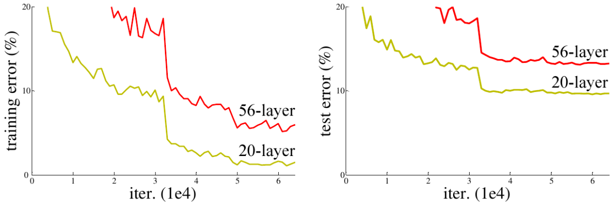
# Problem

- With an increase of the depth of a network, a *degradation* problem arises, were accuracy gets saturated, and then degrades rapidly.

# Problem

- With an increase of the depth of a network, a *degradation* problem arises, were accuracy gets saturated, and then degrades rapidly.

- This is not due to *overfitting*, and adding more layers leads to *higher training error*.



# Solution

**Deep residual learning framework** - Instead of directly learning a desired underlying mapping $\mathcal{H}(\mathbf{x})$, a *residual* mapping $\mathcal{F}(\mathbf{x})$is learned instead, were $\mathcal{F}(\mathbf{x}) := \mathcal{H}(\mathbf{x}) - \mathbf{x}$.

# Solution

**Deep residual learning framework** - Instead of directly learning a desired underlying mapping $\mathcal{H}(\mathbf{x})$, a *residual* mapping $\mathcal{F}(\mathbf{x})$is learned instead, were $\mathcal{F}(\mathbf{x}) := \mathcal{H}(\mathbf{x}) - \mathbf{x}$.

This is done by using `shortcut` connections.