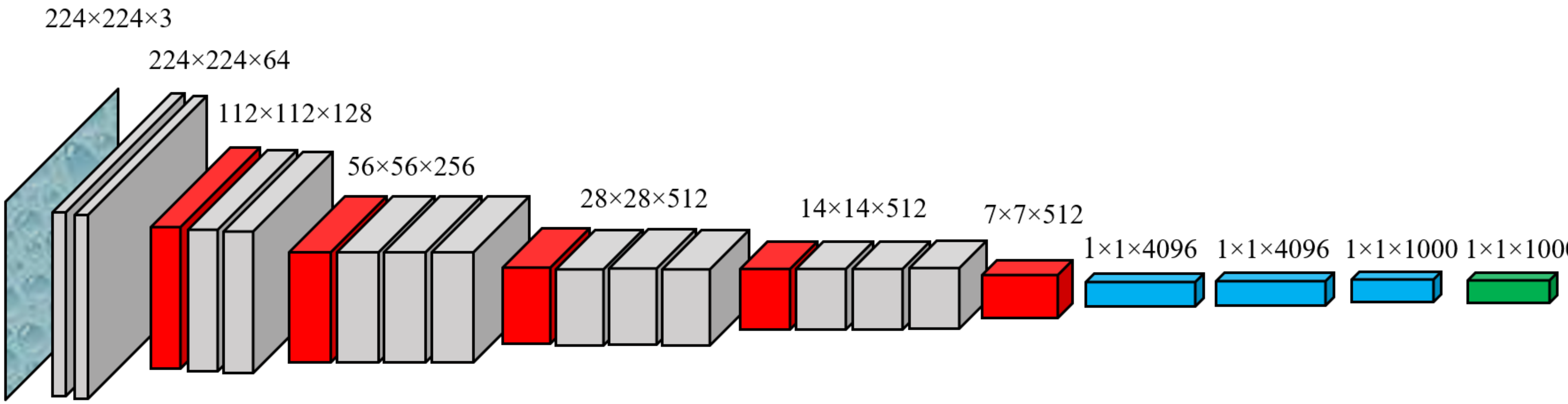
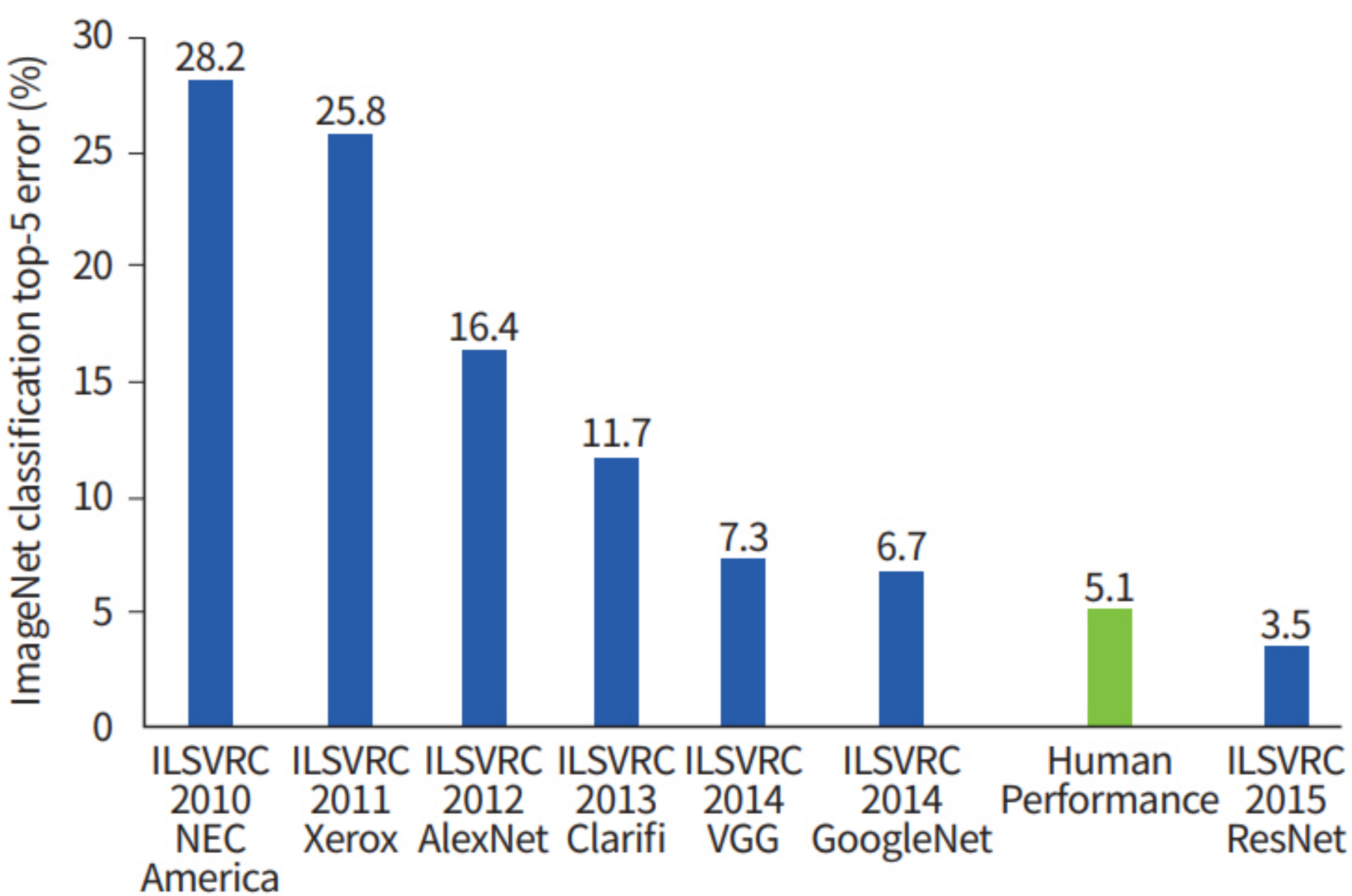
FELLINI GENERAL MEETING

STATISTICAL LEARNING THEORY GEOMETRICALLY STRUCTURED DATA

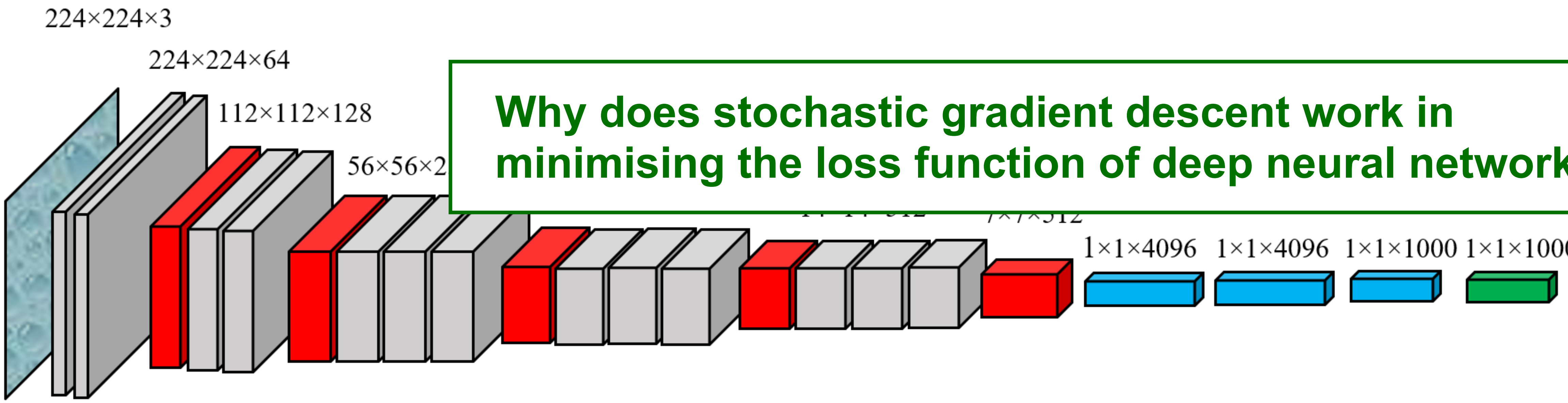
Pietro Rotondo — University of Milan

DEEP LEARNING: SUCCESS IN APPLICATIONS VS LACK OF THEORETICAL UNDERSTANDING



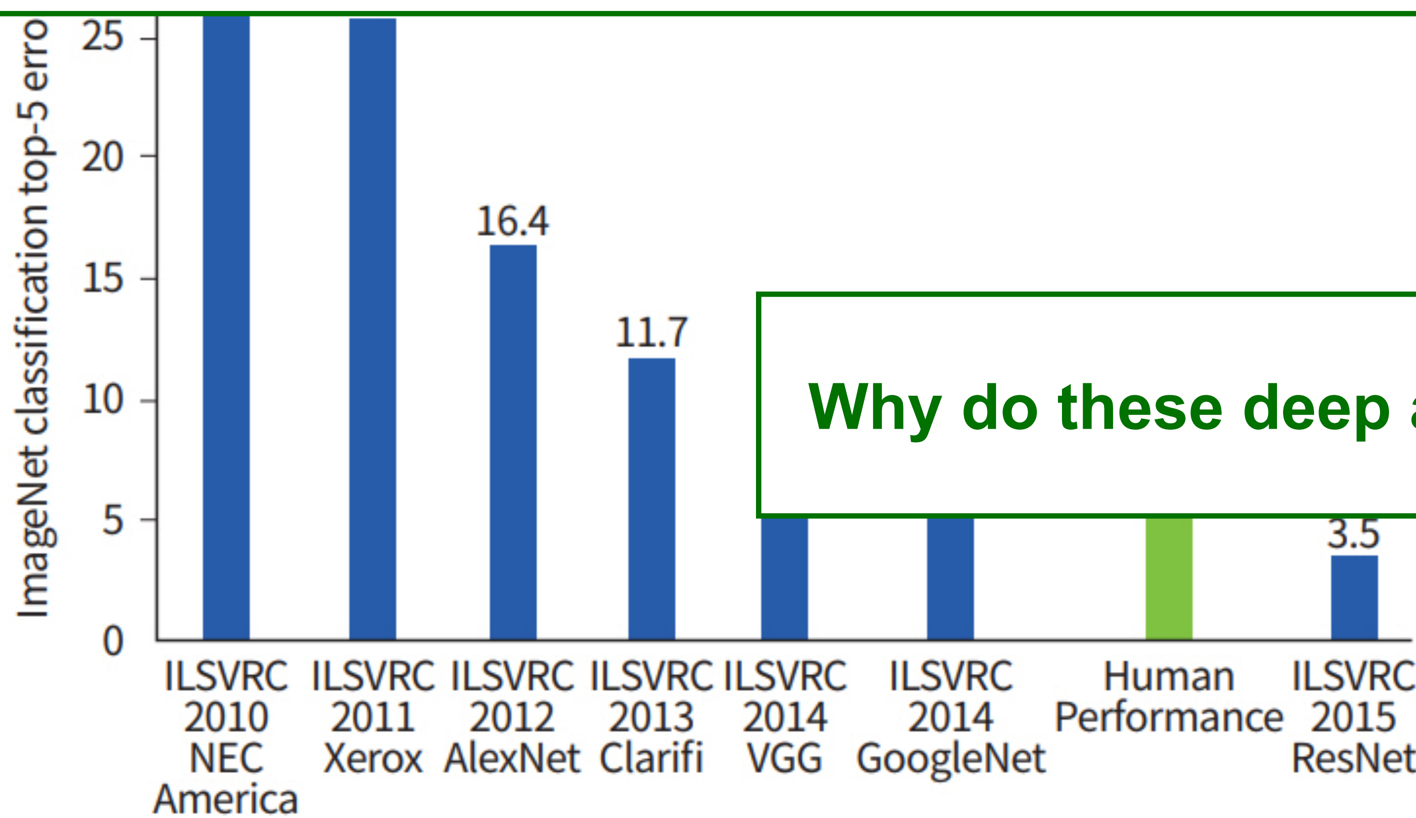


DEEP LEARNING: SUCCESS IN APPLICATIONS VS LACK OF THEORETICAL UNDERSTANDING



Why does stochastic gradient descent work in minimising the loss function of deep neural networks?

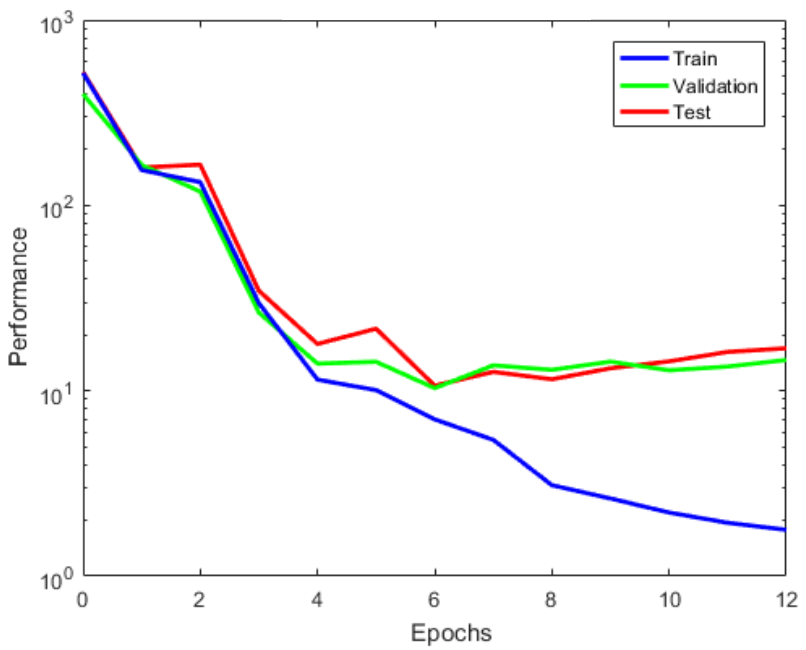
Why do we need depth and we cannot obtain the same performance with a shallow network?



Why do these deep architectures generalise well?

NEITHER STATISTICAL PHYSICS AND THEORETICAL COMPUTER SCIENCE PROVIDE SATISFACTORY ANSWERS

GENERALIZATION = the ability of a given architecture to perform well on unseen data



- **STATISTICAL PHYSICS** does provide the generalization performance in the so-called "Teacher-Student scenario"

  PROS exact results in the thermodynamic limit!

  CONS results hold only for very simple architectures (perceptron, committee machine, Support vector machine)

- **COMPUTER SCIENCE** (and Statistical Learning Theory in particular) offer upper bounds to the generalization performance

  PROS bounds are universal, they do not depend on the architecture at hand

  CONS bounds are too loose to be useful in practice

INTERDISCIPLINARY RESEARCH EFFORTS MAY DELIVER NOVEL INSIGHTS ON THE PROBLEM

- **CONSIDERING MORE REALISTIC MODELS OF DATA STRUCTURE**

  Perceptual manifolds inspired by Neuroscience (H. Sompolinsky et al.)

  Hidden manifold models (L. Zdeborova et al.)

  This is what I have done in my first year as a FF

- **PROVIDING ANALYTICALLY TRACTABLE LIMITS** of Deep Neural Networks

  Mainly the so-called **N**eural **T**angent **K**ernel limit
  (introduced by Computer Scientists in 2018)

  This is what I would like to look more in detail in the next years