# Research Seminar in Data Science
# Paper Reviews

Christoph Schmidl

s4226887

c.schmidl@student.ru.nl

April 18, 2018

**Reviewed papers:**

- **Understanding Deep Learning Requires Rethinking Generalization** by Chiyuan Zhang et al. (Presented by Franka Buytenhuijs)

- **Evolution Strategies as a Scalable Alternative to Reinforcement Learning** by Tim Salimans et al. (Presented by Matthijs Biondina)

# 1 Understanding Deep Learning Required Rethinking Generalization

## 1.1 Summary

In his paper "Understanding Deep Learning Required Rethinking Generalization", Chiyuan Zhang et al state that convolutional neural networks are able to be trained on a dataset with a zero training error as long as the number of paramters execeeds the number of datapoints. This property is descibed as effective capacity. Standard regularization techniques like data augmentation, l2-regularization (weight decay) and dropout do not seem to be essential for a neural network to generalize well but rather serves the purpose as a last optimization step on a neural network architecture which generalizes already well enough based on its inherent architecture. The authors mention that this statement also holds when the underlying relation of the used dataset is exchanged by random labels as long as the number of tunable parameters is high enough and exceeds the amount of datapoints.

## 1.2 Evidence

The authors cite 31 different research papers which seem to be valuable and important for the argumentation. The authors give references to different regularization techniques like l2, dropout, batch-normalization where they decide beween explicit and implicit techniques. The execution of the experiments is backed up by literature and they are using different complexity measures which seem to be plausable in using like VC dimension, Rademacher complexity and uniform stability.

## 1.3 Strengths

The paper is written in a clear way and it was easy to follow. The experiments seems plausible.

## 1.4 Weaknesses

I expected a clear answer to the research question and I got the impression that it was not answered sufficiently just by saying that this inherent regularization property remains a mystery.

## 1.5 Evaluation

I guess this paper would get accepted to a conference or journal because it adds details on top of existing knowledge. However, it does not propose new techniques it just examines the inherent property of neural networks of implicit regularizers.

## 1.6 Comments on the quality of the writing

The writing style was formal but also written in a clear way. I did not find any typos.

## 1.7 Queries for discussion

- Is this knowledge applicable to real-world application or is it just investigating an inherent but interesting property of neural networks which remains a mystery?

- Is this paper also relevant in regards to optimization techniques other than SGD?

# 2 Evolution Strategies as a Scalable Alternative to Reinforcement Learning

## 2.1 Summary

Tim Salimans et al describe the different strengts of Evolution Strategies in comparison to standard techniques which are based on the calculation of gradients. The authors describe the general idea behind black-box optimization, Evolution Strategies and Natural Evolution Strategies where each one is a subset of the other one, ordered from highest to lowest. Unique properties of ES seem to make it a viable approach in real-life application, like invariant to action frequency and delayed rewards, tolerant of extremely long horizons and it does not need temporal discounting. The authors describe the use of ES in the context of hyperparameter tuning of neural networks.

## 2.2 Evidence

The authors cite 45 different research papers where the oldest is from 1973 and goes until 2016. Biological inspired algorithms are quite old and the authors are referencing papers which get a new drive by applying todays computation capabilities. The paper is well backed up and the authors seem to cite a wide variety of sources.

## 2.3 Strengths

The authors describe the general idea behind behind evolution strategies and go into great detail. They are describing that evolution strategies should be more used today because they have inherent properties which might make them a better choice in comparison to gradient based approaches. I worked with biological inspired algorithms in the past and I find it interesting that they did not lose they purpose besides that fact that deep learning approaches seem to take over. It seems that they can be used in conjunction.

## 2.4 Weaknesses

The proposed experiments are often used in the evaluation of biological inspired algorithms and they seem to perform well. But I'd rather seen experiments which are not often used in this context. On the other hand, this would make this paper not comparable to prior work.

## 2.5 Evaluation

This paper would probably be accepted to a conference because it revives evolution strategies in the context of tuning neural networks. It is well backed up by prior work and it contains great detail in the argumentation.

## 2.6 Comments on the quality of the writing

I think that the quality of writing and overall structure is pretty well and the technical detail is also on a high level.

## 2.7 Queries for discussion

- The paper is referencing the usage of CPU threads rather than GPU. I assume that it is also applicable to GPU. So, is this approach used in practice on GPU instances? Are there real life applications?

- Is this paper state-of-the-art or are there other biological inspired approach which outperform natural evolution strategies?