Dynamic Sparse Training: challenges and opportunities towards scalable, efficient, and sustainable AI

Decebal Constantin Mocanu
Associate Professor, Department of Computer Science, University of Luxembourg
Guest Faculty Member, Department of Mathematics and Computer Science, Eindhoven University of Technology

11 APRIL 2023

SMART DIASPORA 2023, TRUST-AI, TIMISOARA

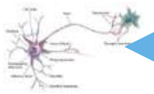
OUTLINE

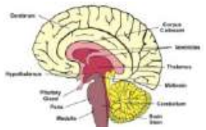
- *Context&Motivation*
- Dense-to-sparse training
- Sparse-to-sparse training
    - principles
    - utilization

SCIENCE PARADIGMS

SCIENCE PARADIGMS

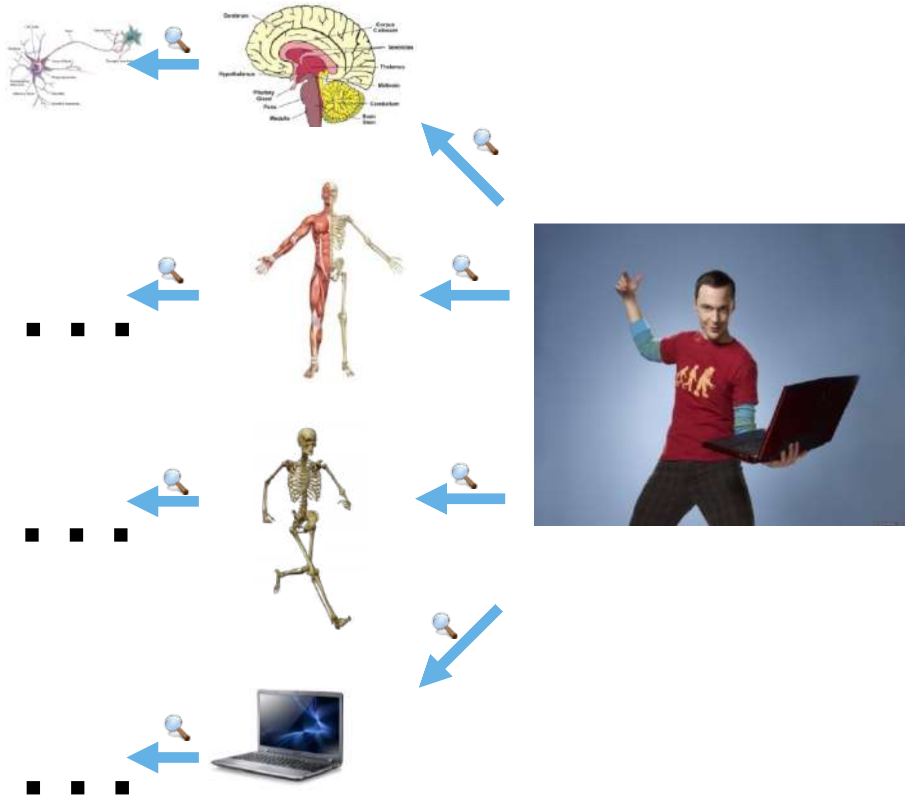
Reductionism





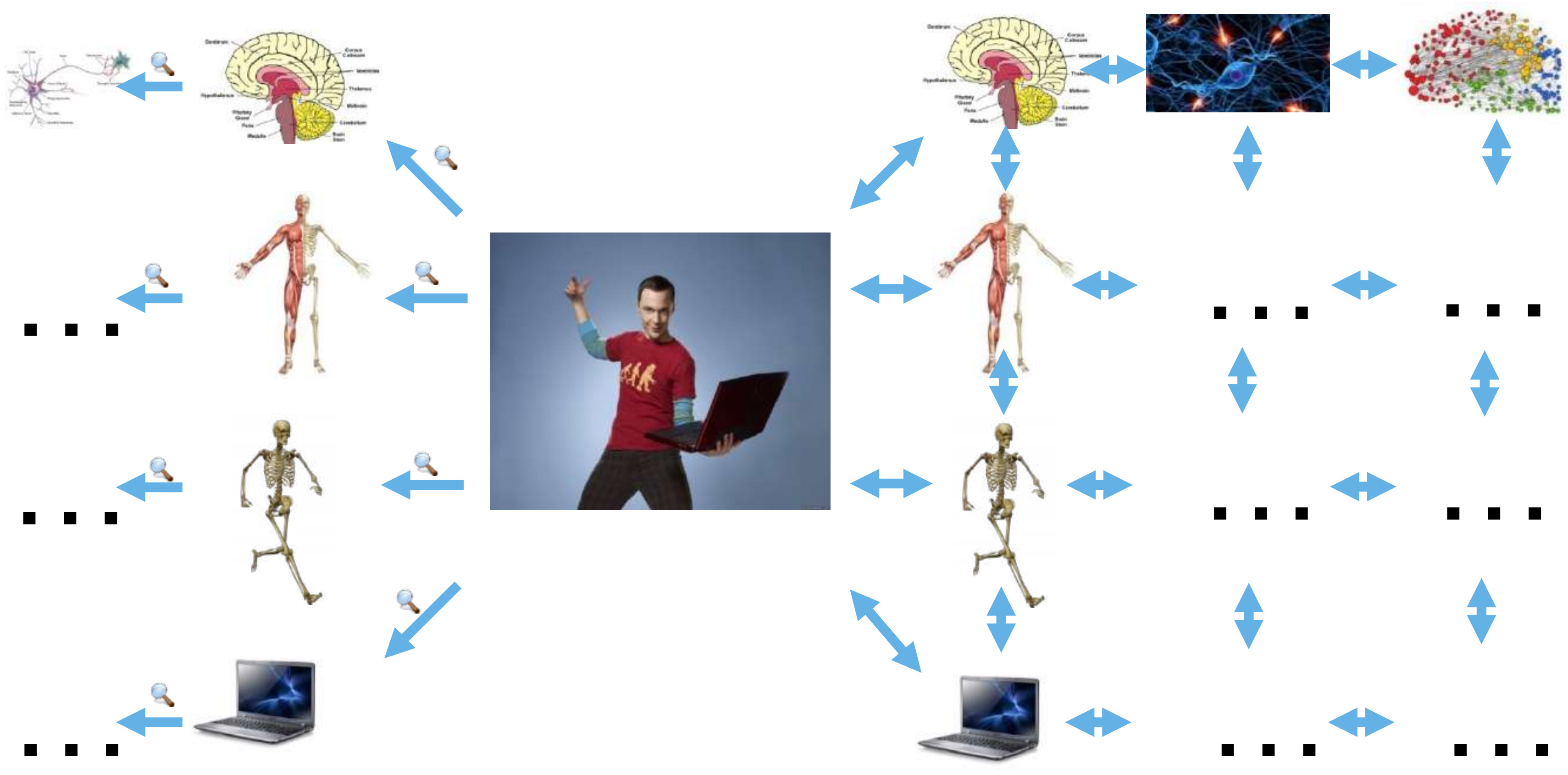
SCIENCE PARADIGMS

Reductionism



SCIENCE PARADIGMS

Reductionism　　　　　　　　　　　*Complex Systems*

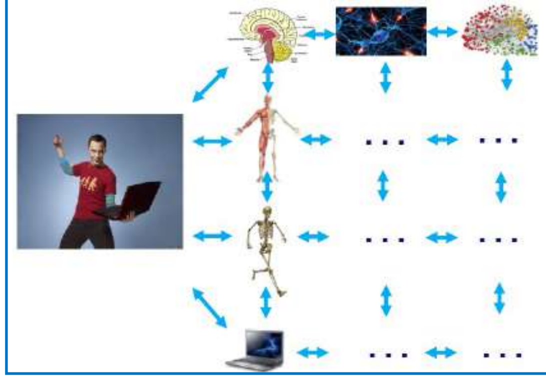


"the totality is not, as it were, a mere heap, but the whole is something beside the parts", Aristotle, **Metaphysics**, ≈350 BC

decebal.mocanu@uni.lu

MY RESEARCH IN A NUTSHELL

Complex systems



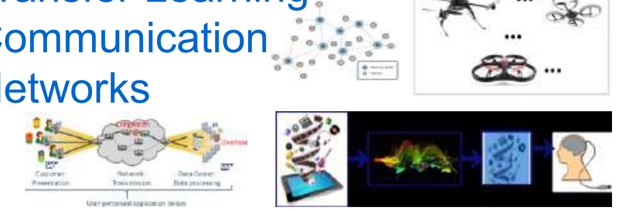
Network Science

Artificial Intelligence

Further Research, e.g.:
- ANN with billions of neurons
- ANN on low-resource devices
- Climate change

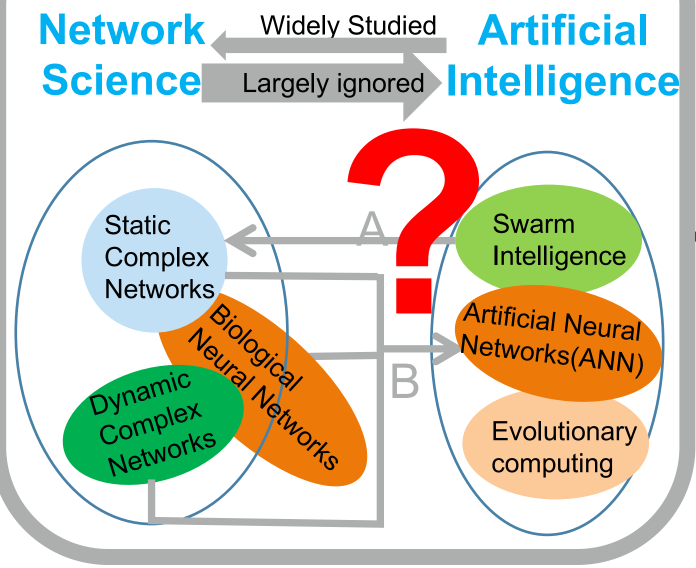
Real-world open questions

- Computer Vision
- Pattern Recognition
- Signal Processing
- Computer Security
- Smart Grid
- Transfer Learning
- Communication Networks



Ongoing and further research, e.g.:
- Continual learning
- Reinforcement learning
- Few-shot learning
- Decentralized learning
- Efficient deep learning
- Sparse training theory

The synergy between



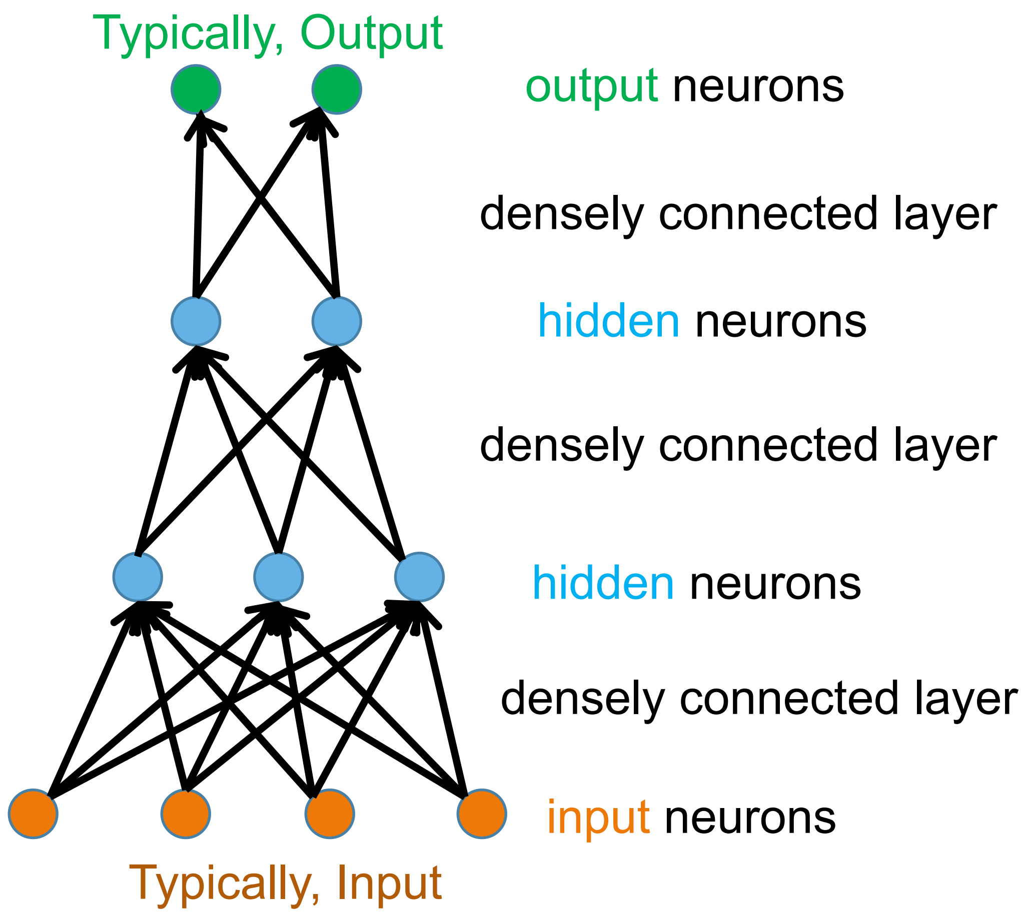
Scalability issues at some levels

References:
- D.C. Mocanu, *On the synergy of network science and artificial intelligence*, **IJCAI** 2016
- D.C. Mocanu, *Network computations in artificial intelligence*, **PhD thesis**, 2017
- [A] D.C. Mocanu, G. Exarchakos, A. Liotta, *Decentralized dynamic understanding of hidden relations in complex networks*, **Scientific Reports**, 2018
- [B] D.C. Mocanu, E. Mocanu, P. Stone, P.H. Nguyen, M. Gibescu, A. Liotta, *Scalable Training of Artificial Neural Networks with Adaptive Sparse Connectivity inspired by Network Science*, **Nature Communications**, 2018

decebal.mocanu@uni.lu

ARTIFICIAL NEURAL NETWORKS (ANN) BRIEF TECHNICAL BACKGROUND

Example of an ANN model
i.e., Multi Layer Perceptron (MLP)



"This model will be a simplification and an idealization, and consequently a falsification", Alan Turing, 1952.

State-of-the-art (broadly speaking):
- ANN models are very good at single specialized tasks (e.g., recognizing cats) -> deep learning hype

RESEARCH AIMS

Long term: study the synergy between network science and artificial intelligence for the advancement of science and societal benefits.

Medium term:

- conceive scalable and efficient artificial neural networks
- capable to perform continual (reinforcement) learning
- and to generalize from limited multimodal data/experiences
- on heterogeneous known and unknown hierarchical tasks

Selected challenges for today talk:

1. **too many connections** to optimize in ANNs (due to densely connected layers); existing software and hardware is optimized just for dense neural networks - limiting their representational power, etc.
2. **catastrophic forgetting** (when learning new knowledge, an ANN model is prone to forget previously learnt knowledge)
3. and many other parts of a very large **puzzle** would have to be solved …

MOTIVATION

Example of real-world problems

1200 pixels

1600 pixels 3 RGB channels

1600x1200x3=5.760.000 dimensions

Biology:
- p53 Mutants Data Set
  - 5409 attributes
- Thrombin dataset
  - 139351 binary attributes
- Human chromosomes
  - Millions of attributes

Low-resources devices:

Extremely High Energy Consumption:
- For instance, GPT-3
  - 175 billion parameters
  - 355 GPU-years for training
  - 4.600.000 USD for training

  [Chuan Li, Lamdalabs, 2020]

State-of-the-art successful solutions and limitations

Deep Convolutional Neural Networks
- Work usually for images
- Usually still need some densely connected layers

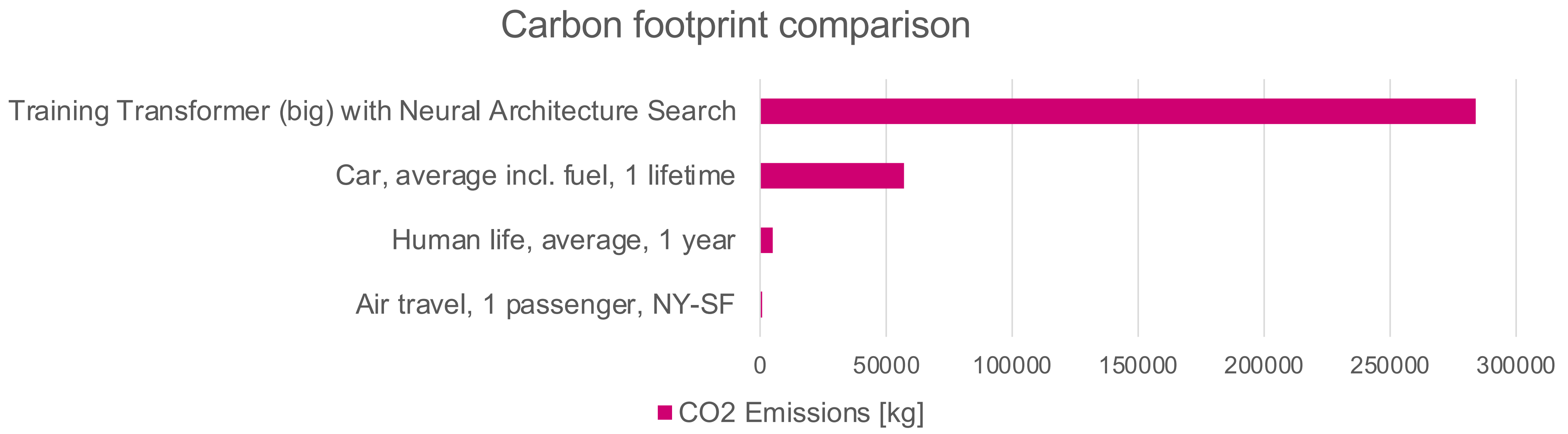
Feature extraction/selection
- Needs expert knowledge of the field
- It is hard to automatize

Pruning methods – model compression
- They still need large densely connected layers
- Extremely high training time
- Models have to be trained in the cloud and after that have to be moved to the device

- *No solution (in practice) for the training phase*
- Pruning for the exploitation phase

How to obtain scalable, performant, and efficient (deep) artificial neural networks?
- To handle better high dimensional data
- For faster computational time in training and exploitation (inference)

MOTIVATION

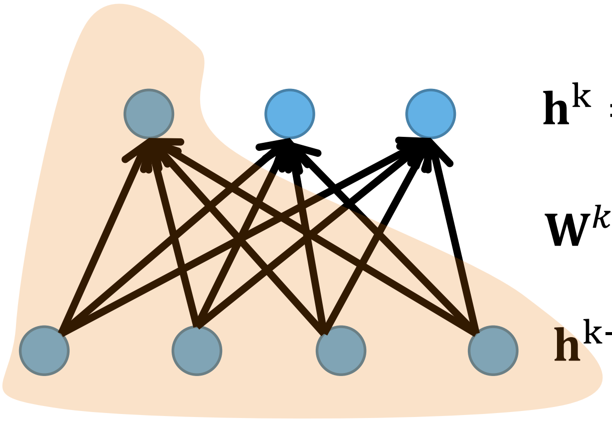


Reference:
- Emma Strubell, Ananya Ganesh, Andrew McCallum, *Energy and Policy Considerations for Deep Learning in NLP,* **ACL** 2019

BACKGROUND

Unnecessary connections in neural networks fully connected layers (problem)

Neural networks Fully-Connected (FC) k layer



$$\mathbf{h}^k = [h_1^k, h_2^k, \dots, h_{n^k}^k]$$

$$\boldsymbol{W}^k \in \boldsymbol{R}^{n^{k-1} \times n^k}$$

$$\mathbf{h}^{k-1} = [h_1^{k-1}, h_2^{k-1}, \dots, h_{n^{k-1}}^{k-1}]$$

Example of weights matrix (W^k) after training

$$\boldsymbol{W}^k = \begin{pmatrix} \boxed{\begin{matrix} 2 \\ -1 \\ 0.02 \\ 0.01 \end{matrix}} & \begin{matrix} 0.001 \\ -0.03 \\ -3 \\ 1.4 \end{matrix} & \begin{matrix} 3 \\ 0.01 \\ -0.02 \\ 0.024 \end{matrix} \end{pmatrix}$$

Weights close to zero don't count too much

$$h_1^k = f(h_1^{k-1}w_{11}^k + h_2^{k-1}w_{21}^k + h_3^{k-1}w_{31}^k + h_4^{k-1}w_{41}^k)$$

Problems with dense weights matrix:

- High computational time to optimize unnecessary weights
- High storage requirements to store unnecessary weights
- Unnecessary multiplications and additions

decebal.mocanu@uni.lu

OUTLINE

- Context&Motivation
- *Dense-to-sparse training*
- Sparse-to-sparse training
  - principles
  - utilization

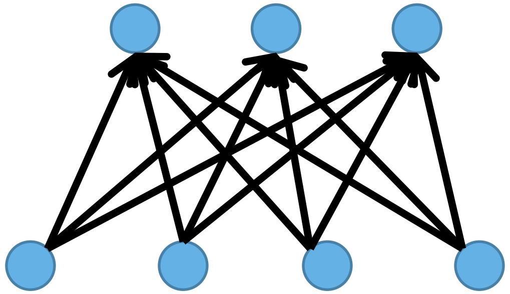
PRUNING

Schematic Procedure:

- *Step1*: Start with a fully-connected neural network
- *Step2*: Train the network until some convergence criteria is satisfied
- *Step3*: Identify unimportant connections based on some saliency criteria. The most used in practice criteria are:
  - Magnitude (connections with values closed to zero)
  - Various information-theoretic criteria
- *Step4*: Prune those unimportant connections
- *Step5*: Repeat from *Step2* until the neural network satisfies the target trade-off sparsity/accuracy

PRUNING - A VISUAL EXAMPLE

Train the neural network until convergence is satisfied.

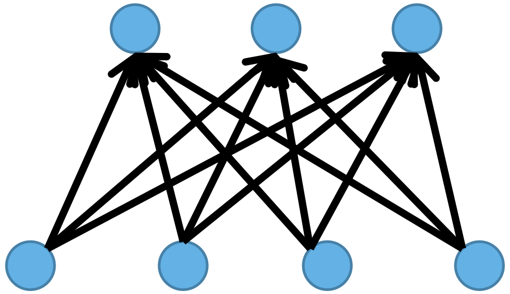


W^k after training

$$\mathbf{W}^k = \begin{pmatrix} 2 & 0.001 & 3 \\ -1 & -0.03 & 0.01 \\ 0.02 & -3 & -0.02 \\ 0.01 & 1.4 & 0.024 \end{pmatrix}$$

PRUNING - A VISUAL EXAMPLE

Train the neural network until convergence is satisfied.



W^k after training

$$\mathbf{W}^k = \begin{pmatrix} 2 & 0.001 & 3 \\ -1 & -0.03 & 0.01 \\ 0.02 & -3 & -0.02 \\ 0.01 & 1.4 & 0.024 \end{pmatrix}$$

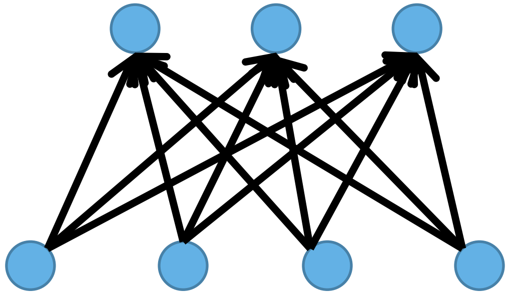
Use magnitude to identify unimportant connections.

decebal.mocanu@uni.lu

PRUNING - A VISUAL EXAMPLE

Train the neural network until convergence is satisfied.

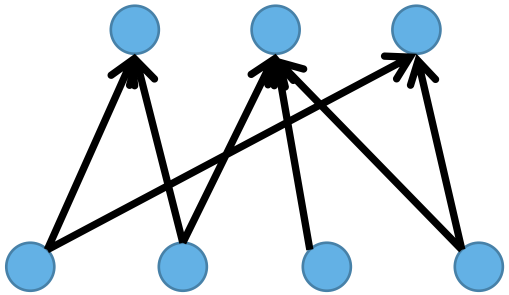
W^k after training



$$\mathbf{W}^k = \begin{pmatrix} 2 & 0.001 & 3 \\ -1 & -0.03 & 0.01 \\ 0.02 & -3 & -0.02 \\ 0.01 & 1.4 & 0.024 \end{pmatrix}$$

Use magnitude to identify unimportant connections.

W^k after pruning



$$\mathbf{W}^k = \begin{pmatrix} 2 & 0 & 3 \\ -1 & -0.03 & 0 \\ 0 & -3 & 0 \\ 0 & 1.4 & 0.024 \end{pmatrix}$$

Prune the unimportant connections.

Repeat until the neural network satisfies the target trade-off sparsity/accuracy.

decebal.mocanu@uni.lu

REFERENCES

There is a considerable amount of work in the last years on this topic. Some are using magnitude and some are using various information metric criteria. It is almost impossible to cover all of it, and for this we apologize to many authors. Below are some references for further read.

Selected References:

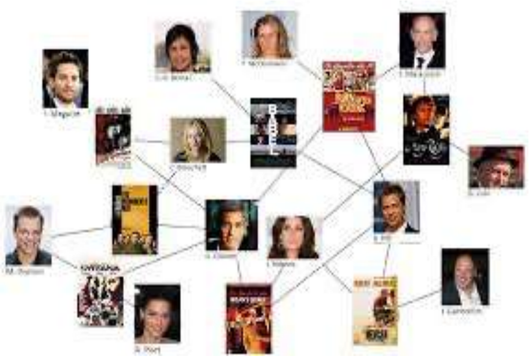
- Michael Mozer & Paul Smolensky. *Using Relevance to Reduce Network Size Automatically*. **Connection Science**, 1989.
- Yan Le Cun, John S. Denker, Sara A. Solla. *Optimal brain damage.* **NIPS**, 1990
- Babak Hassibi, David G. Stork. *Second order derivatives for network pruning: Optimal Brain Surgeon.* **NIPS**, 1993
- Song Han, Jeff Pool, John Tran, William J. Dally. *Learning both Weights and Connections for Efficient,* **NIPS**, 2015
- Dmitry Molchanov, Arsenii Ashukha, Dmitry Vetrov. *Variational Dropout Sparsifies Deep Neural Networks,* **ICML**, 2017

OUTLINE

- Context&Motivation
- Dense-to-sparse training
- *Sparse-to-sparse training*
  - *principles*
  - utilization

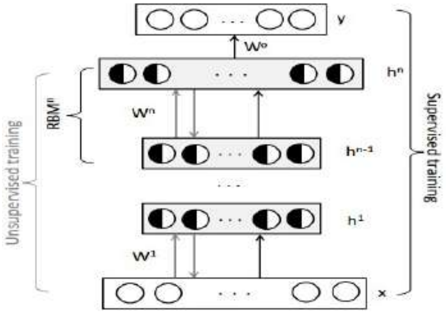
INSPIRATION

Social Networks

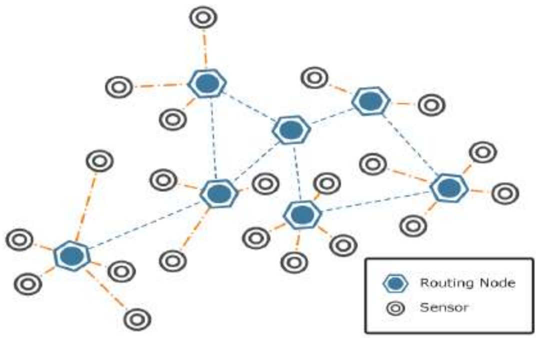


Transportation Networks

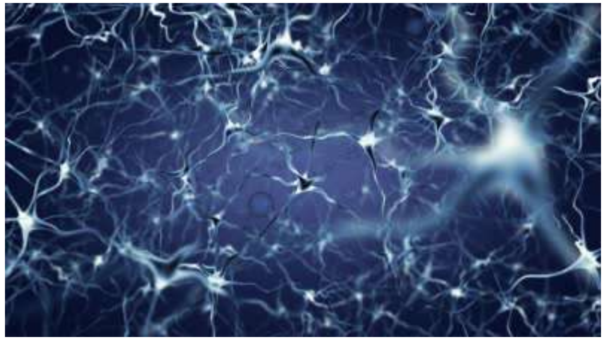
Artificial Neural Networks



Wireless Sensors Networks

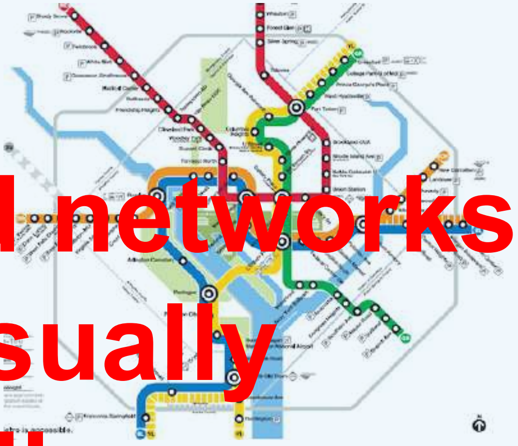


Biological Neural Networks



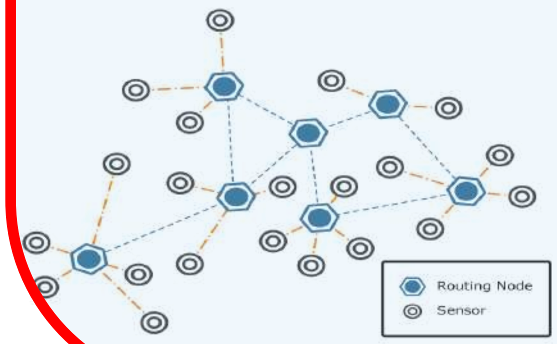
INSPIRATION

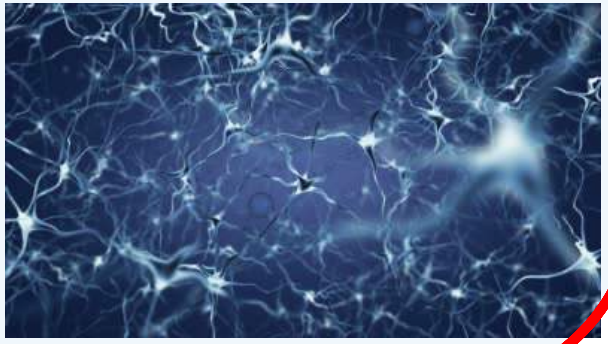
Social Networks | **Transportation Networks**



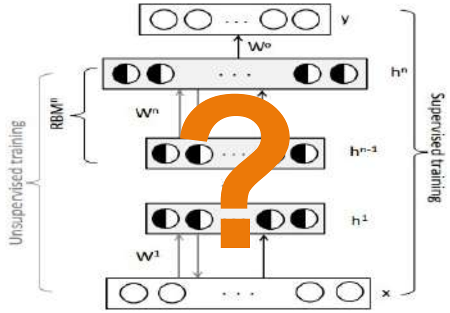
Real world networks are usually intrinsically sparse and evolve in time!

Wireless Sensors Networks | **Biological Neural Networks**





Artificial Neural Networks



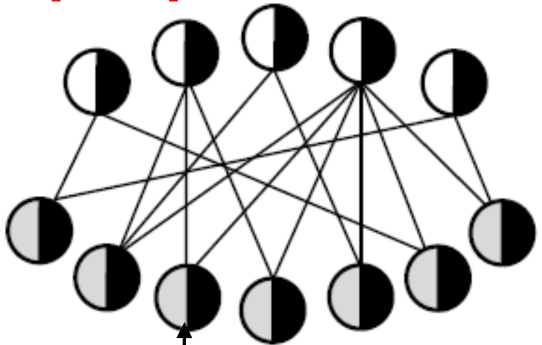
PIECE 1 OF THE PUZZLE – PROPOSED CONCEPT
ALWAYS SPARSE – NEVER DENSE

Our statements:

- ANN **sparsity must be** initiated from the design phase, **before training**!!!!

On par performance with dense training!

Data Statistics

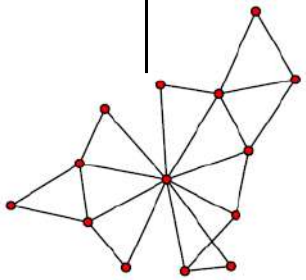


XBM

- sparse model
- static sparse connectivity
- fixed before training

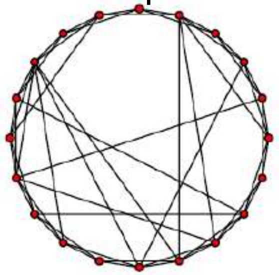
References:

- D.C. Mocanu, E. Mocanu, P. Nguyen, M. Gibescu, A. Liotta, *A topological insight into restricted Boltzmann machines*, **Machine Learning Journal**, ECML-PKDD, 2016.



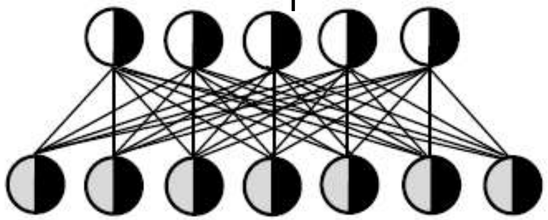
Scale-free topology

+



Small-world topology

+



Restricted Boltzmann machines

decebal.mocanu@uni.lu

PIECE 1 OF THE PUZZLE – PROPOSED CONCEPT ALWAYS SPARSE – NEVER DENSE

Our statements:

- ANN **sparsity** must be initiated from the design phase, **before training**!!!!
- While training with static sparse connectivity leads to decent enough performance, adaptive (dynamic) sparse connectivity allows the model to properly fit the data and, consequently, to ensure higher performance!!!!

Adaptive sparse connectivity **goal**:

- Both, the connectivity graph and the connection weights, are optimized together during training!

References:

- D.C. Mocanu, E. Mocanu, P. Nguyen, M. Gibescu, A. Liotta, *A topological insight into restricted Boltzmann machines*, **Machine Learning Journal**, ECML-PKDD, 2016.
- D.C. Mocanu, *Network computations in artificial intelligence*, **PhD thesis**, 2017
- D.C. Mocanu, E. Mocanu, P. Stone, P.H. Nguyen, M. Gibescu, A. Liotta, *Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science*, arXiv:1707.04780, **Nature Communications, 2018**

PIECE 1 OF THE PUZZLE – PROPOSED SOLUTION
SPARSE EVOLUTIONARY TRAINING (SET)

SET – an efficient algorithm to train sparsely connected layers for ANNs with adaptive sparse connectivity which can safely replace any densely connected layer, at no loss in accuracy, while having quadratically less parameters. Thus:

- quadratically faster training time
- quadratically smaller memory footprint

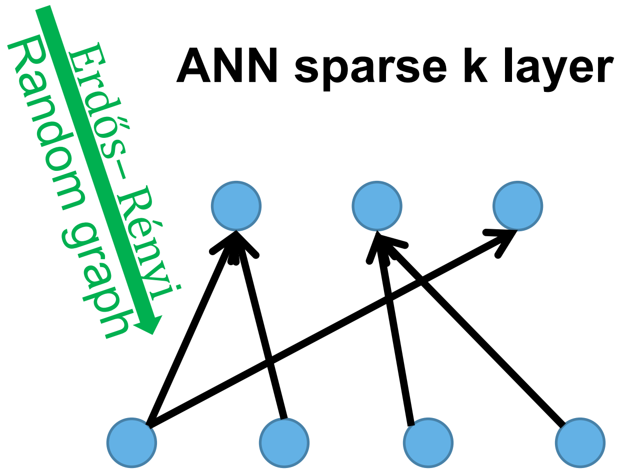
References:

- D.C. Mocanu, *Network computations in artificial intelligence*, **PhD thesis**, 2017
- D.C. Mocanu, E. Mocanu, P. Stone, P.H. Nguyen, M. Gibescu, A. Liotta: *Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science*, arXiv:1707.04780, **Nature Communications**, 2018

SET **procedure:**

- start with an Erdős–Rényi random graph topology for any bipartite layer:
  $p(\mathrm{w}_{\mathrm{ij}}^{\mathrm{k}}) = \varepsilon(\mathrm{n}^{\mathrm{k}} + \mathrm{n}^{\mathrm{k}+\mathbf{1}})/(\mathrm{n}^{\mathrm{k}}\mathrm{n}^{\mathrm{k}+\mathbf{1}})$; ε – small integer, control sparsity level
- for each training epoch
  - perform standard training procedure;
  - for each bipartite layer:
    - remove a fraction ζ of the smallest positive weights;
    - remove a fraction ζ of the largest negative weights;
    - add randomly new weights (connections) in the same amount as the ones removed previously

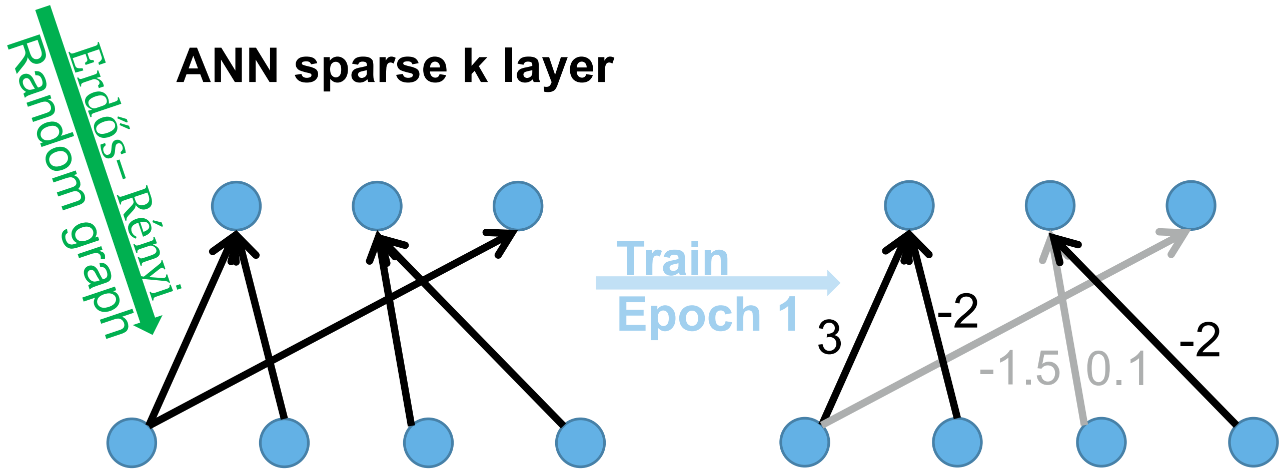
PIECE 1 OF THE PUZZLE – MECHANISM
SPARSE EVOLUTIONARY TRAINING (SET)



https://github.com/dcmocanu/sparse-evolutionary-artificial-neural-networks (201 stars)

PIECE 1 OF THE PUZZLE – MECHANISM
SPARSE EVOLUTIONARY TRAINING (SET)

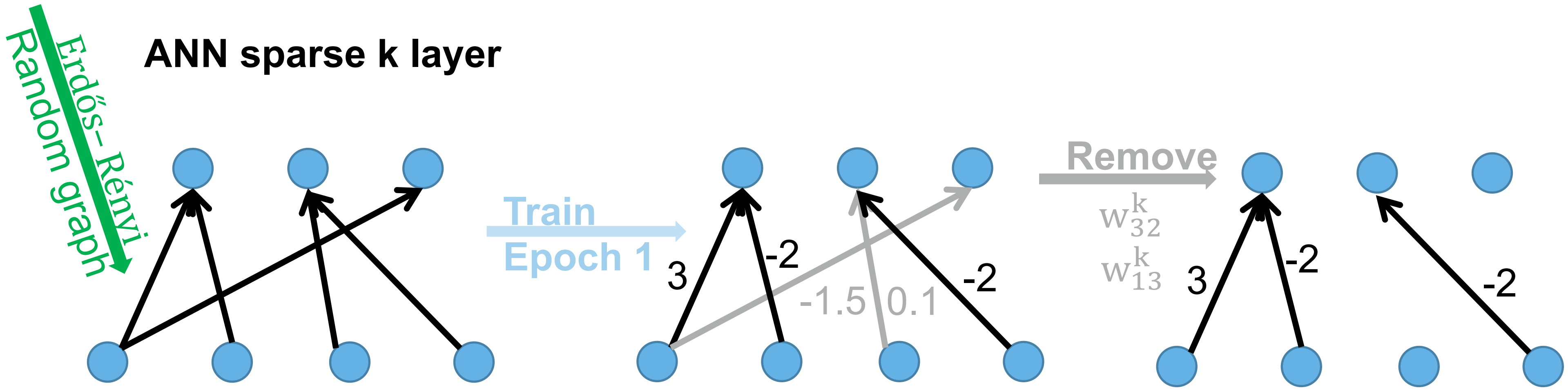
ANN sparse k layer



https://github.com/dcmocanu/sparse-evolutionary-artificial-neural-networks (201 stars)

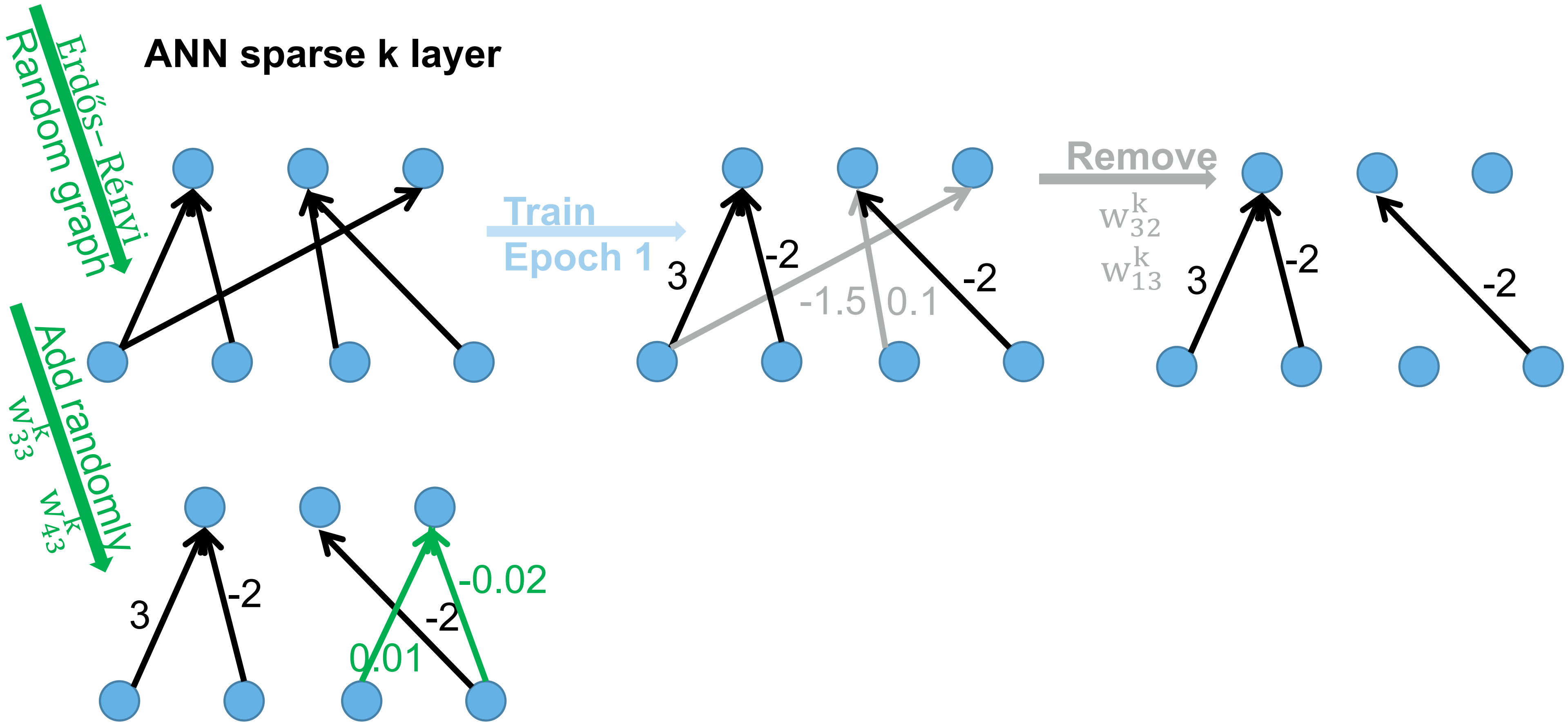
PIECE 1 OF THE PUZZLE – MECHANISM
SPARSE EVOLUTIONARY TRAINING (SET)

ANN sparse k layer



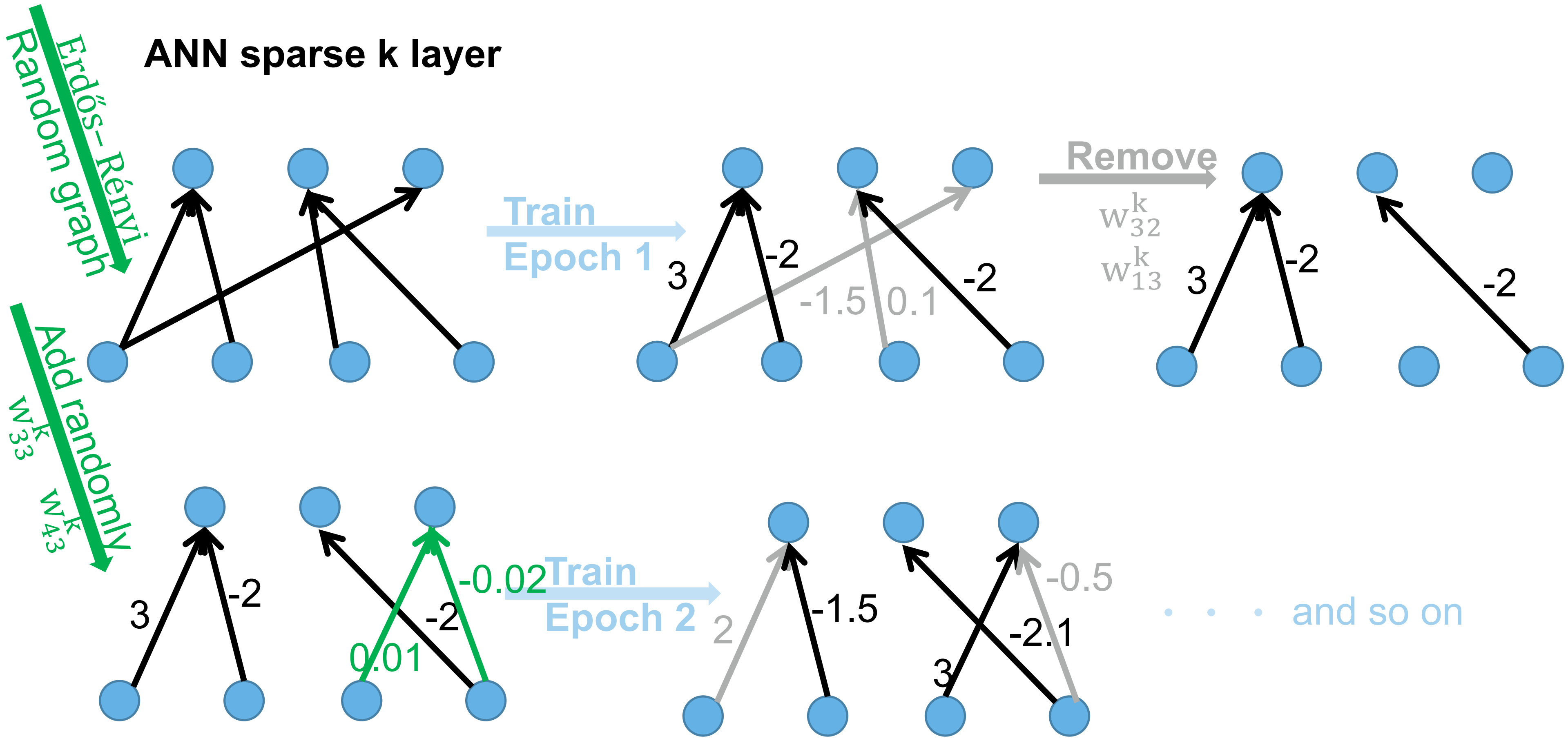
https://github.com/dcmocanu/sparse-evolutionary-artificial-neural-networks (201 stars)

PIECE 1 OF THE PUZZLE – MECHANISM
SPARSE EVOLUTIONARY TRAINING (SET)



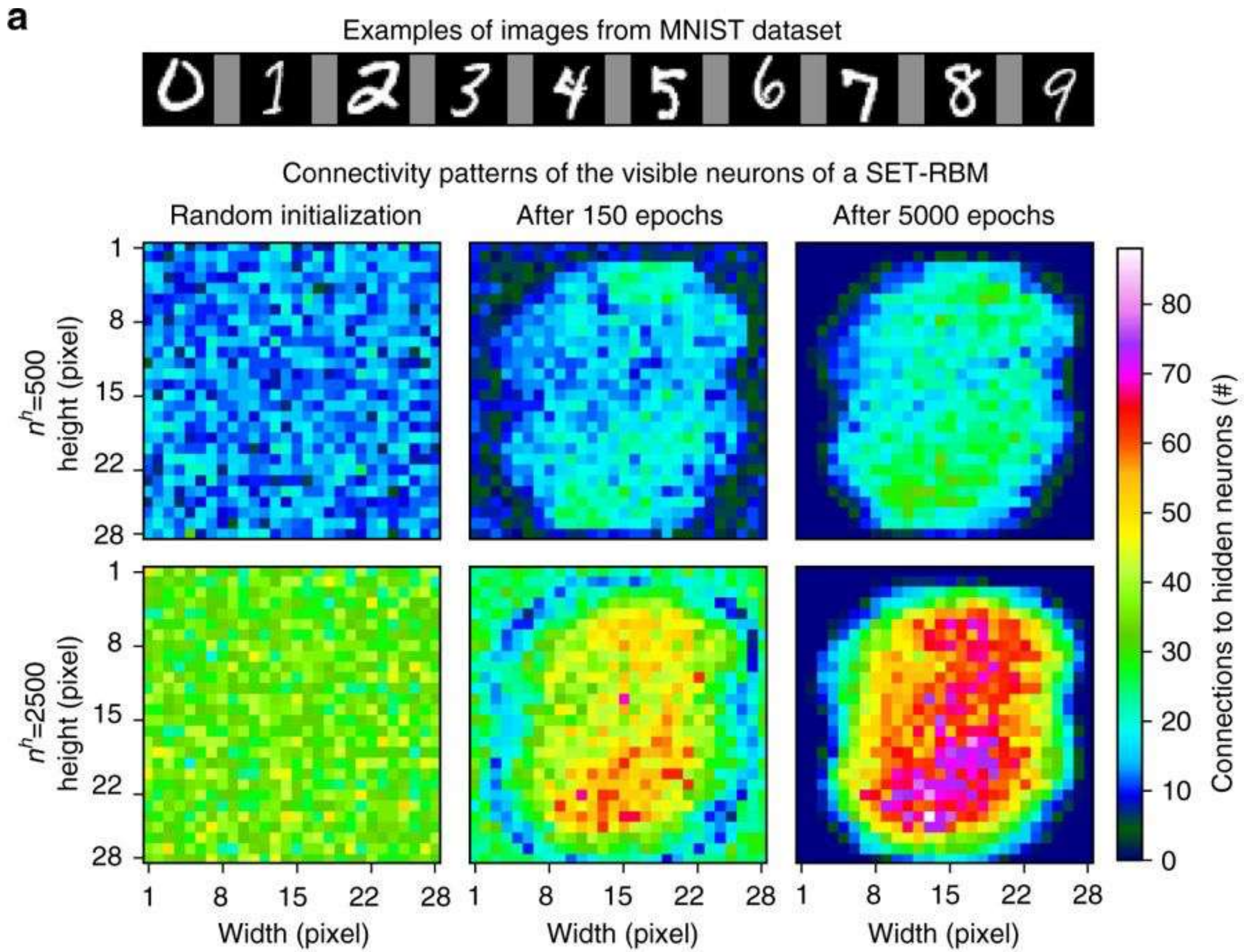
https://github.com/dcmocanu/sparse-evolutionary-artificial-neural-networks (201 stars)

PIECE 1 OF THE PUZZLE – MECHANISM
SPARSE EVOLUTIONARY TRAINING (SET)



https://github.com/dcmocanu/sparse-evolutionary-artificial-neural-networks (201 stars)

PIECE 1 OF THE PUZZLE
SET – AUTOMATIC PATTERN DISCOVERY

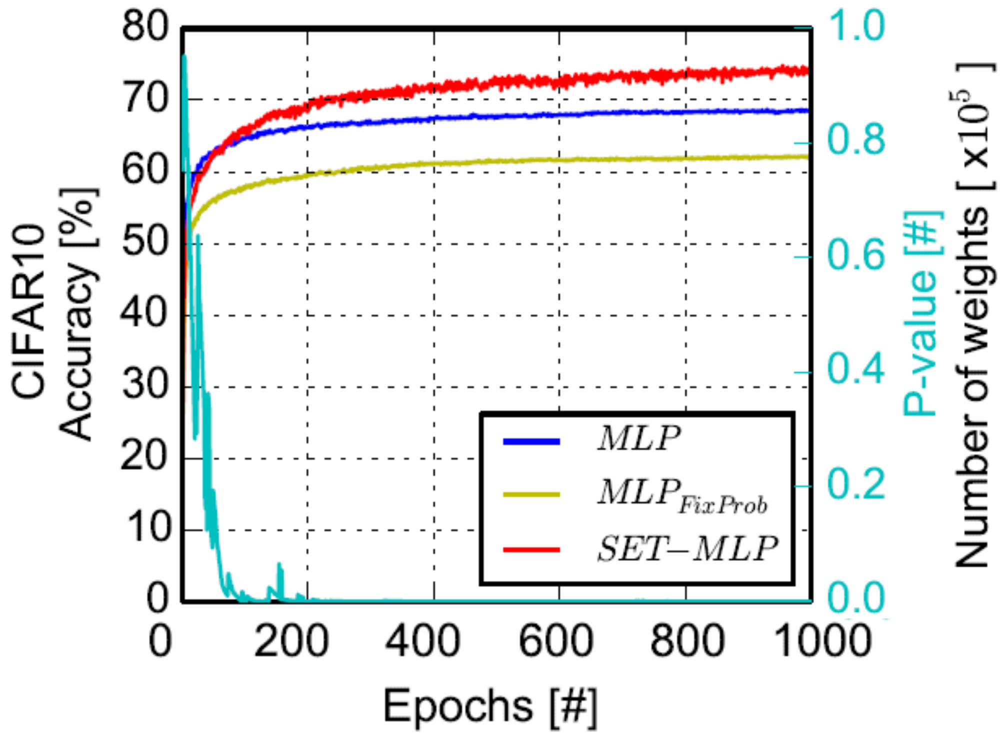


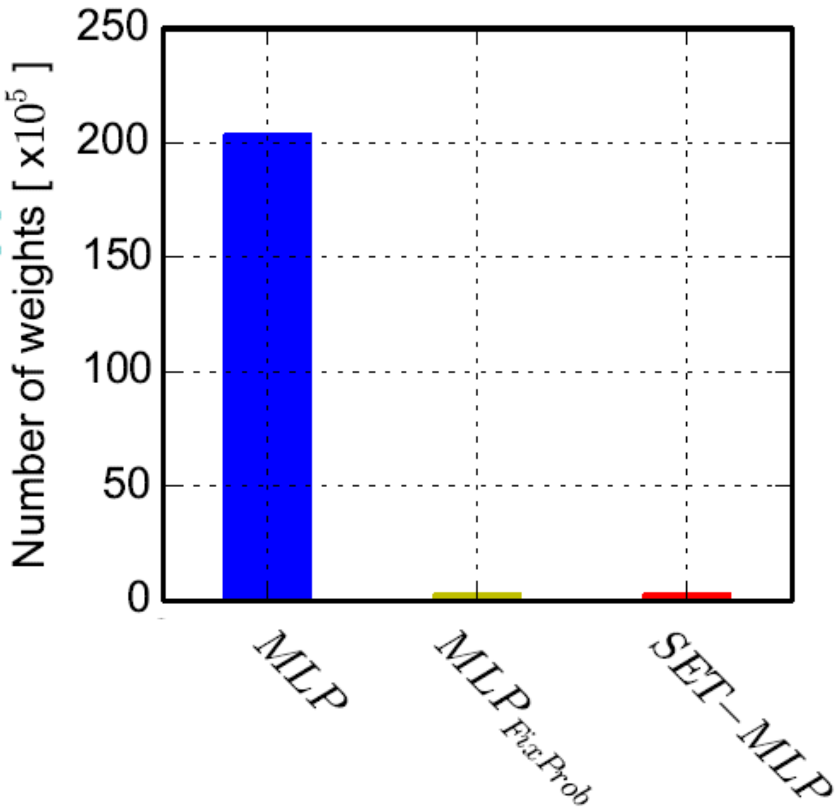
References:

- D.C. Mocanu, E. Mocanu, P. Stone, P.H. Nguyen, M. Gibescu, A. Liotta, *Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science*, arXiv:1707.04780, **Nature Communications**, 2018

PIECE 1 OF THE PUZZLE

SET – SUPERVISED LEARNING





Reference:

- D.C. Mocanu, E. Mocanu, P. Stone, P.H. Nguyen, M. Gibescu, A. Liotta, *Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science*, arXiv:1707.04780, **Nature Communications**, 2018

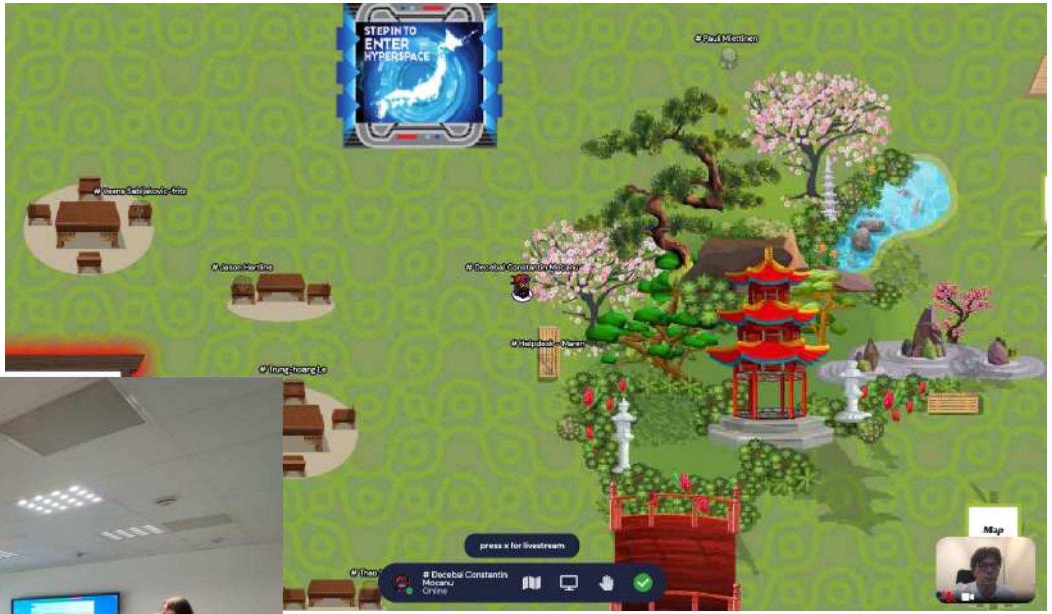
Dataset:

- CIFAR 10 (object recognition)
- 10 classes

PIECE 1 OF THE PUZZLE – INTERMEZZO

Impact: SET (Sparse Evolutionary Training) put the basis of the emerging subfield of **(Dynamic) Sparse Training.**

D.C. Mocanu, E. Mocanu, Z. Vale, D. Ernst, *Scalable Deep Learning*
(T24), tutorial at **International Joint Conference on Artificial Intel**



canu, T. Pinto, Z. Vale, *Scalable Deep Learning: how far is one billion*
ternational Joint Conference on Artificial Intelligence (IJCAI), 2020.

S. Liu, G. Sokar, Z. Atashgahi, D.C. Mocanu, E. Mocanu, *Sparse Neural Networks Training*, tutorial at **ECMLPKDD** 2022 ☺

…pandemic times …

decebal.mocanu@uni.lu

OUTLINE

- Context&Motivation
- Dense-to-sparse training
- *Sparse-to-sparse training*
    - principles
    - *utilization*

PIECE 1 OF THE PUZZLE
TRULY SPARSE IMPLEMENTATIONS
ONE MILLION NEURONS ON A LAPTOP WITHOUT GPU

Microarray gene expression data:
- very high dimensionality
- small number of samples
- e.g., Leukemia dataset (54,675 features, 2096 samples, 18 classes)
- traditional approach (feature selection/extraction + classifier)

Reference:
- Shiwei Liu, Decebal Constantin Mocanu, Amarsagar Reddy Ramapuram Matavalam, Yulong Pei, Mykola Pechenizkiy, *Sparse evolutionary Deep Learning with over one million artificial neurons on commodity hardware*, arXiv:1901.09181, **Neural Computing and Applications**, 2021

Model	Hardware	Density level (%)	Total time (s) (train + test)	Accuracy (%)
Small SET-MLP	1 CPU thread	1.04	65	82.88 ± 1.18
Large SET-MLP	1 CPU thread	0.007	4914	81.83 ± 1.11
mrPSVM [40]	Conventional	n/a	1265	81.1
mrPSVM [40]	Hadoop cluster	n/a	291	81.1

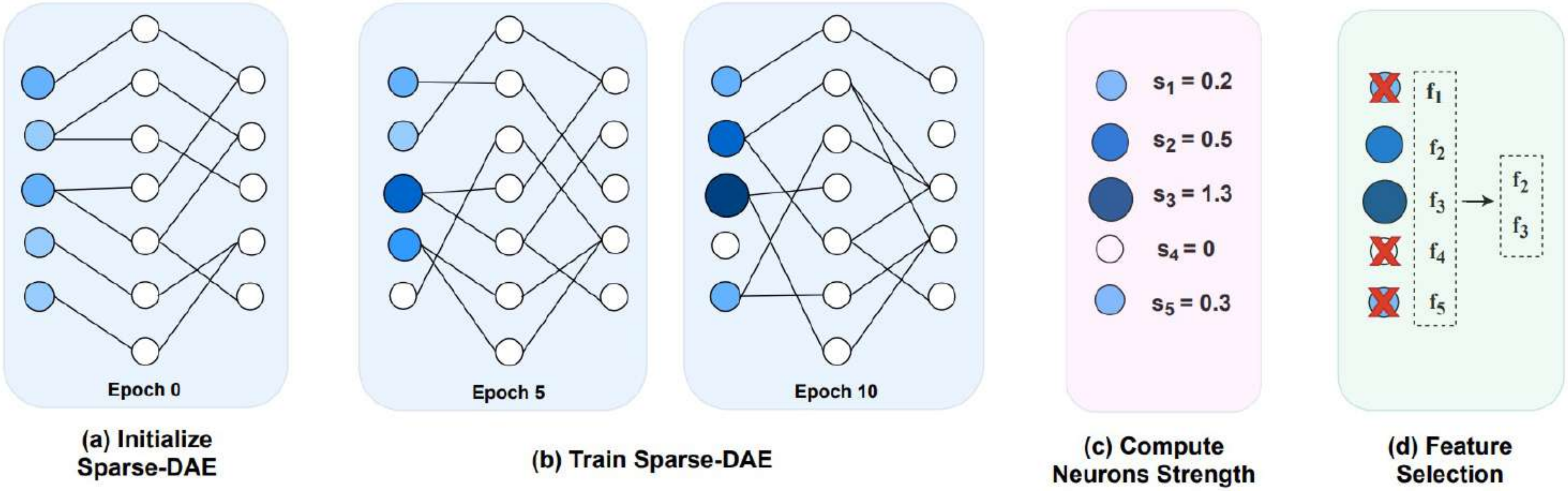
- mrPSVM (ensemble of feature selection algorithms + support vector machine) from Kumar M, Rath SK (2015) *Classification of microarray using mapreduce based proximal support vector machine classifier*, **Knowl Based Syst** 89:584–602

decebal.mocanu@uni.lu

PIECE 1 OF THE PUZZLE

QUICKSELECTION (QS)

DYNAMIC SPARSE TRAINING FOR UNSUPERVISED FEATURE SELECTION



DAE: Denoising Autoencoder

Neuron Strength: $s_i = \sum_{j=1}^{n^1} |W_{ij}^1|$

References:

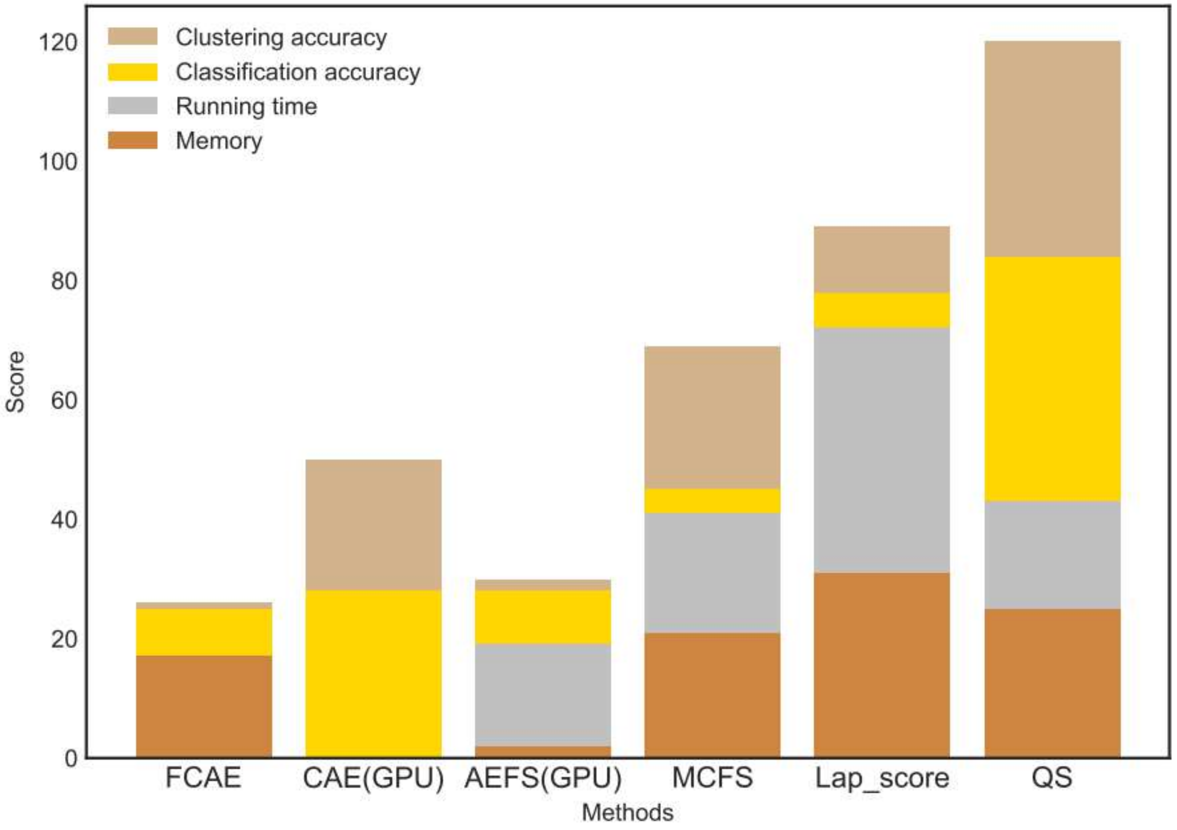
- Zahra Atashgahi, Ghada Sokar, Tim van der Lee, Elena Mocanu, Decebal Constantin Mocanu, Raymond Veldhuis, Mykola Pechenizkiy, *Quick and Robust Feature Selection: the Strength of Energy-efficient Sparse Training for Autoencoders,* **Machine Learning journal**, ECML-PKDD 2022

PIECE 1 OF THE PUZZLE

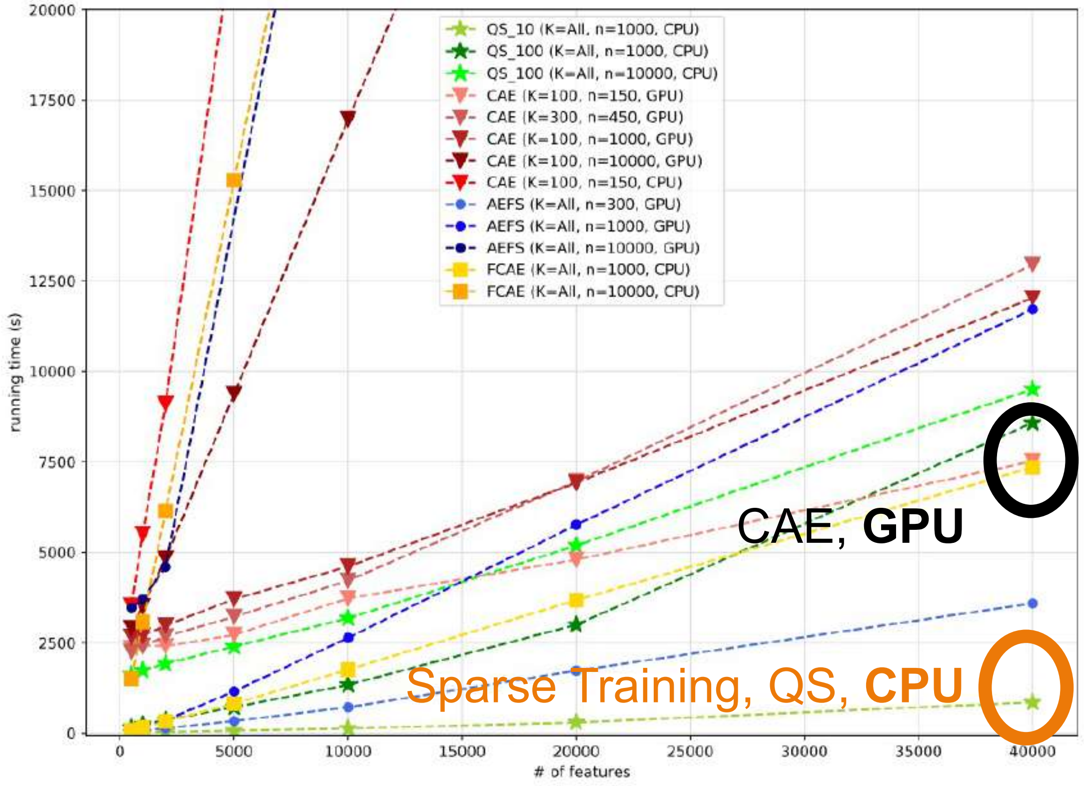
QUICKSELECTION (QS)

DYNAMIC SPARSE TRAINING FOR UNSUPERVISED FEATURE SELECTION

Results summary (over 8 datasets, ranked 1st or 2nd)



Artificial dataset – running time



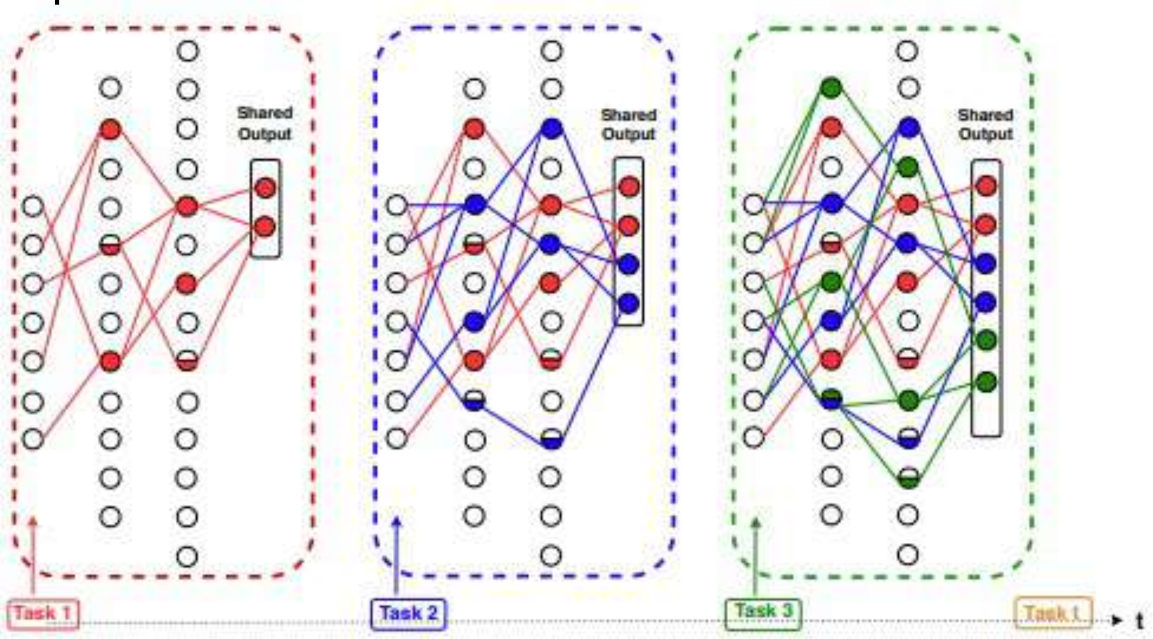
References:

- Zahra Atashgahi, Ghada Sokar, Tim van der Lee, Elena Mocanu, Decebal Constantin Mocanu, Raymond Veldhuis, Mykola Pechenizkiy, *Quick and Robust Feature Selection: the Strength of Energy-efficient Sparse Training for Autoencoders*, **Machine Learning journal**, ECML-PKDD 2022

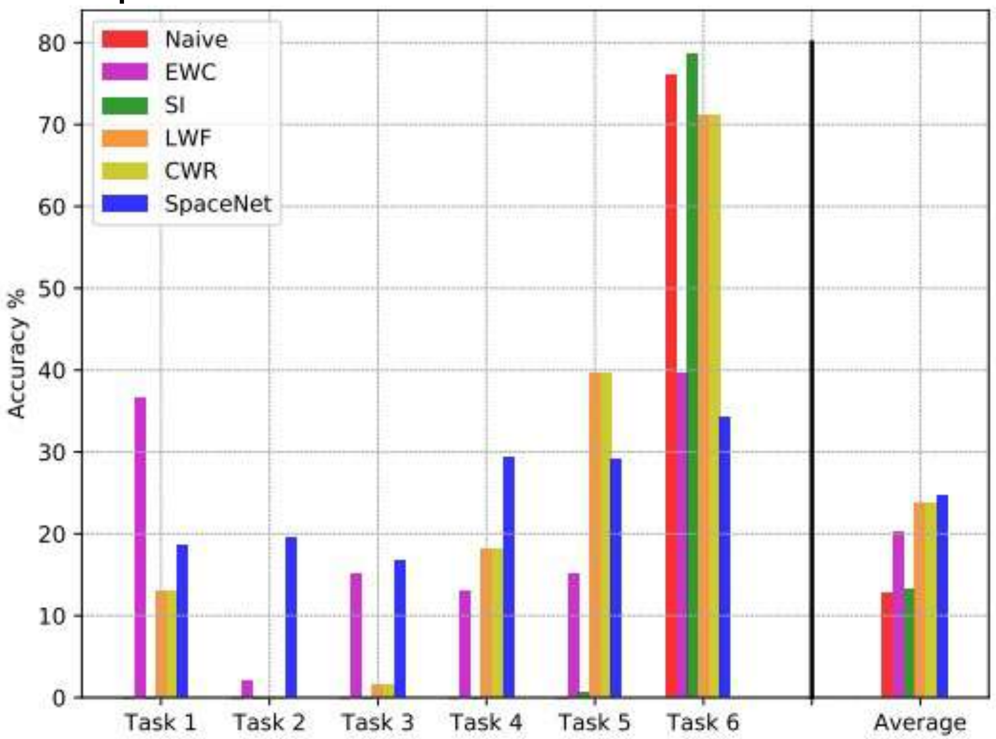
decebal.mocanu@uni.lu

PIECE 1 AND 2 OF THE PUZZLE
SPACENET
SPARSE TRAINING AND CONTINUAL LEARNING

SpaceNet schematic architecture



SpaceNet results on CIFAR 10/100

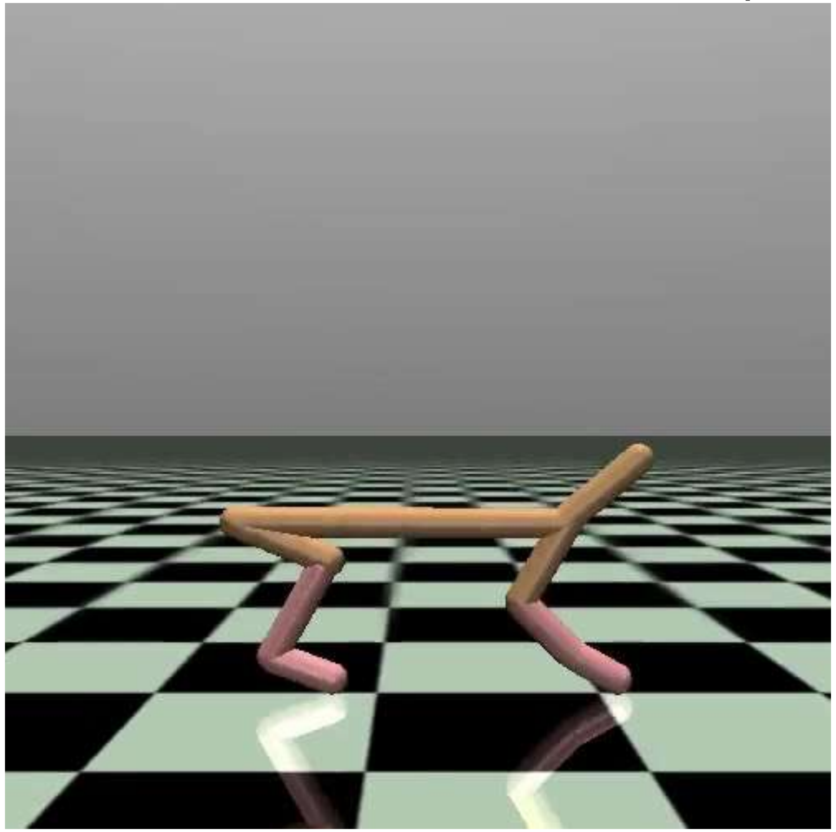


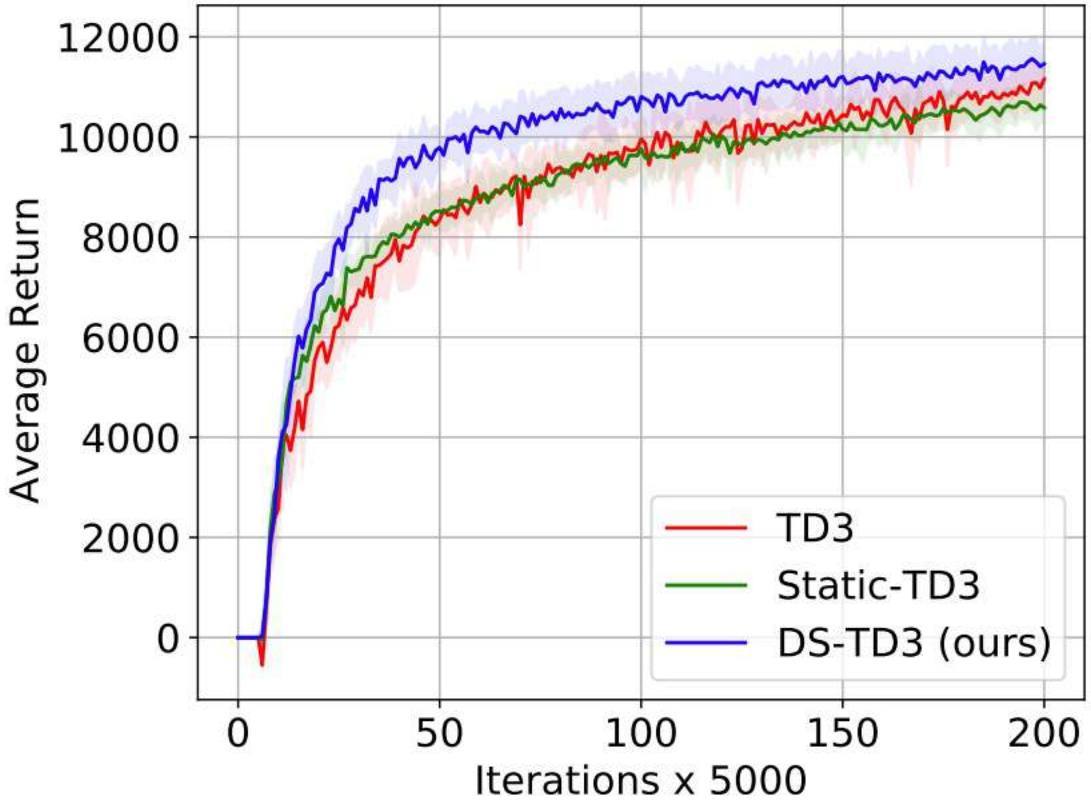
Reference:

- Ghada Sokar, Decebal Mocanu, Mykola Pechenizkiy, *SpaceNet: Make Free Space For Continual Learning*, arXiv:2007.07617 (2020), **Neurocomputing**, 2021

PIECE 1 OF THE PUZZLE
DYNAMIC SPARSE TD3 (DS-TD3)
SPARSE TRAINING AND DEEP REINFORCEMENT LEARNING

Continous control tasks with sparse MLPs optimized by sparse training. An example on Half Cheetah.





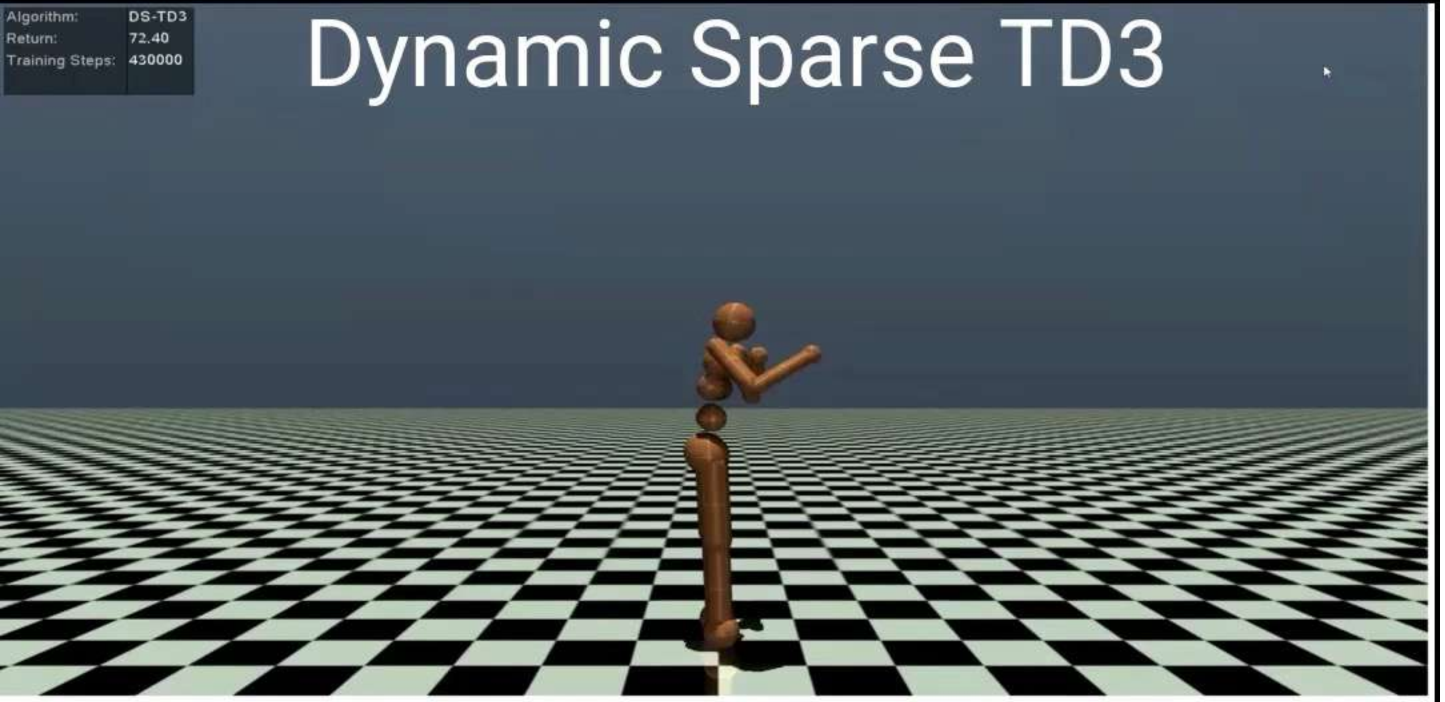
dense connectivity

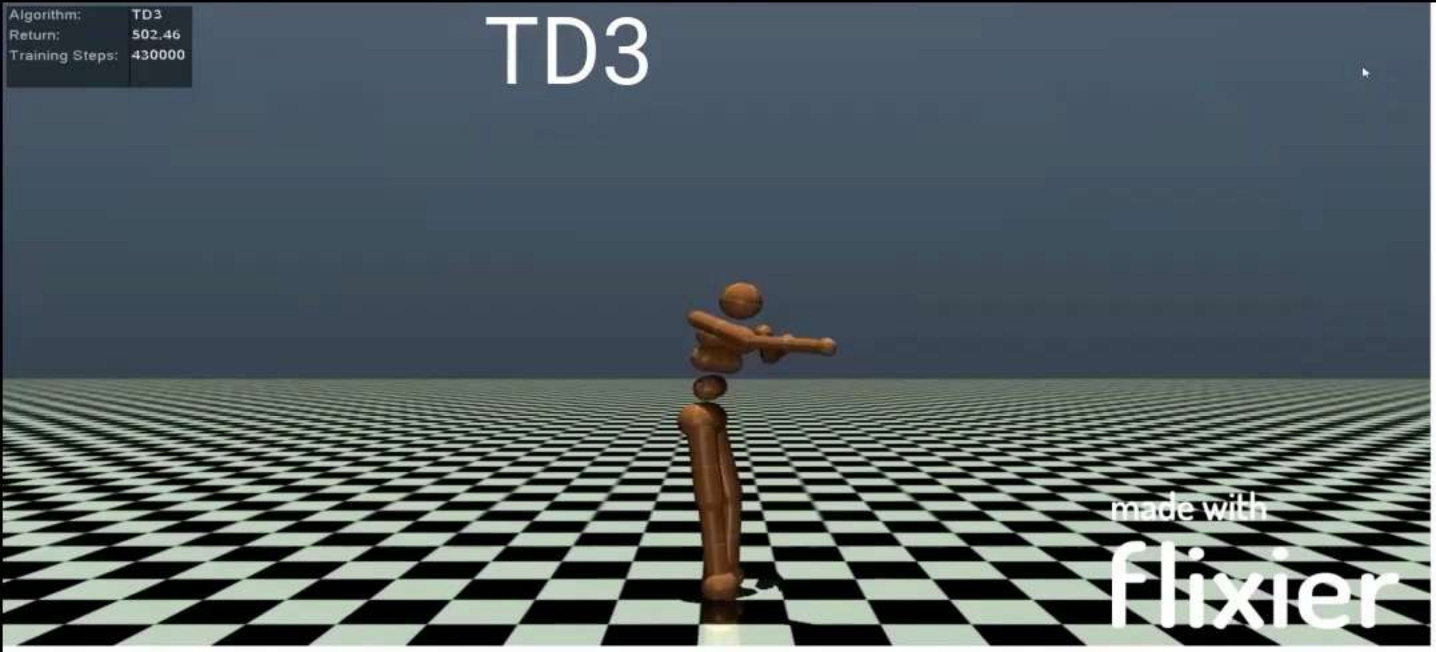
static sparse connectivity

adaptive sparse connectivity
(50% sparsity level)

Reference:

- Ghada Sokar, Elena Mocanu, Decebal Constantin Mocanu, Mykola Pechenizkiy, Peter Stone, *Dynamic sparse training for deep reinforcement learning,* arXiv:2106.04217, **IJCAI** 2022, (and best paper award at AAMAS 2022 - **ALA** workshop)





PIECE 1 OF THE PUZZLE

SPARSE TRAINING AND DEEP REINFORCEMENT LEARNING

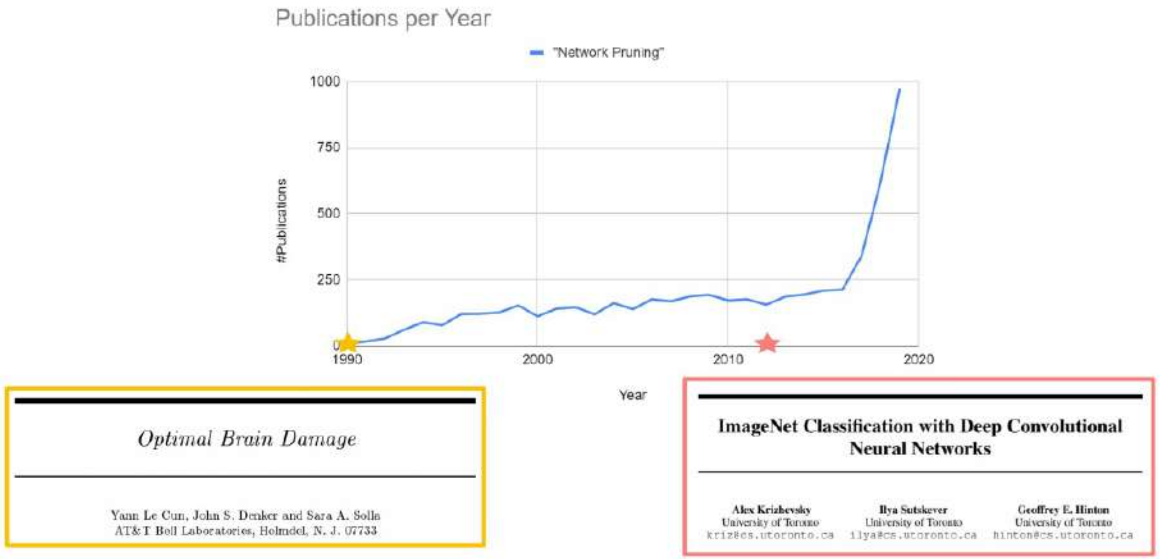
Reference:

- Ghada Sokar, Elena Mocanu, Decebal Constantin Mocanu, Mykola Pechenizkiy, Peter Stone, *Dynamic sparse training for deep reinforcement learning,* arXiv:2106.04217, **IJCAI** 2022, (and best paper award at AAMAS 2022 - **ALA** workshop)

PIECE 1 OF THE PUZZLE
SPARSE TRAINING – CHALLENGES AND FUTURE

- Deep learning hardware and software is optimized for dense matrices (main challenge)
- Sparse training behaviour is not fully understood yet

The situation is changing - NVIDIA A100 GPU (released in 2020) supports 2:4 sparsity



Reference:
- Jeff Pool, *Accelerating sparsity in NVIDIA ampere architecture*, 2020

Would you like to know more?
ICLR SNN 2023 Workshop, 5 May, hybrid, Kigali, Rwanda.
(*Sparsity in Neural Networks: On practical limitations and tradeoffs between sustainability and efficiency*)
https://www.sparseneural.net/

PIECE 1 OF THE PUZZLE
SPARSE TRAINING – SUPPLEMENTARY REFERENCES

Algorithms
- Guillaume Bellec, David Kappel, Wolfgang Maass & Robert Legenstein. *Deep rewiring: training very sparse deep net,* **ICLR** 2018
- Hesham Mostafa, Xin Wang. *Parameter Efficient Training of Deep Convolutional Neural Networks by Dynamic Sparse Reparameterization,* **ICML** 2019
- Siddhant M. Jayakumar, Razvan Pascanu, Jack W. Rae, Simon Osindero, Erich Elsen, *Top-KAST: Top-K Always Sparse Training,* **NeurIPS** 2020
- Jaeho Lee, Sejun Park, Sangwoo Mo, Sungsoo Ahn, Jinwoo Shin, *Layer-adaptive sparsity for the magnitude-based pruning*, **ICLR** 2021
- Selima Curci, Decebal C. Mocanu, Mykola Pechenizkiy*, Truly sparse neural networks at scale,* under review in **Nature Communications** (2nd revision), 2021
- Tianlong Chen, Yu Cheng, Zhe Gan, Lu Yuan, Lei Zhang, Zhangyang Wang, *Chasing Sparsity in Vision Transformers: An End-to-End Exploration,* **NeurIPS** 2021
- Shiwei Liu, Decebal C. Mocanu, Yulong Pei, Mykola Pechenizkiy, *Selfish Sparse RNN Training,* **ICML** 2021
- Shiwei Liu, Tianlong Chen, Xiaohan Chen, Zahra Atashgahi, Lu Yin, Huanyu Kou, Li Shen, Mykola Pechenizkiy, Zhangyang Wang, Decebal C. Mocanu, *Sparse Training via Boosting Pruning Plasticity with Neuroregeneration,* **NeurIPS** 2021
- Geng Yuan et al., *MEST: Accurate and Fast Memory-Economic Sparse Training Framework on the Edge,* **NeurIPS** 2021
- Zahra Atashgahi, Joost Pieterse, Shiwei Liu, Decebal C. Mocanu, Raymond Veldhuis, Mykola Pechenizkiy, *A Brain-inspired Algorithm for Training Highly Sparse Neural Networks*, **Machine Learning** (**ECMLPKDD** 2022 journal track)

Sparse neural networks understanding
- Shiwei Liu, Tim van der Lee, Anil Yaman, Zahra Atashgahi, Davide Ferraro, Ghada Sokar, Mykola Pechenizkiy, Decebal C. Mocanu, *Topological insights into sparse neural networks*, **ECMLPKDD** 2020
- Kale-ab Tessera, Sara Hooker, Benjamin Rosman, *Keep the Gradients Flowing: Using Gradient Flow to Study Sparse Network Optimization, arXiv:2102.01670,* 2021
- Shiwei Liu, Lu Yin, Decebal C. Mocanu, Mykola Pechenizkiy, *Do We Actually Need Dense Over-Parameterization? In-Time Over-Parameterization in Sparse Training,* **ICML** 2021

Applications & customized algorithms
- [Federated Learning] Sameer Bibikar, Haris Vikalo, Zhangyang Wang, Xiaohan Chen. *Federated Dynamic Sparse Training: Computing Less, Communicating Less, Yet Learning Better,* **AAAI** 2022
- **[**Reinforcement Learning] Laura Graesser, Utku Evci, Erich Elsen, Pablo Samuel Castro, *The State of Sparse Training in Deep Reinforcement Learning*, **ICML** 2022
- [Continual Learning] Ghada Sokar, Decebal C. Mocanu, and Mykola Pechenizkiy, *Avoiding Forgetting and Allowing Forward Transfer in Continual Learning via Sparse Networks,* **ECMLPKDD** 2022
- [Feature Selection] Ghada Sokar, Zahra Atashgahi, Mykola Pechenizkiy, Decebal C. Mocanu, *Where to Pay Attention in Sparse Training for Feature Selection?,* **NeurIPS** 2022
- [Time Series Classification] Qiao Xiao, Boqian Wu, Yu Zhang, Shiwei Liu, Mykola Pechenizkiy, Elena Mocanu, Decebal C. Mocanu, *Dynamic Sparse Network for Time Series Classification: Learning What to "See"*, **NeurIPS 2022**

Surveys and open questions
- Torsten Hoefler, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, Alexandra Peste, *Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks*, **JMLR**, 2021
- Decebal C. Mocanu, Elena Mocanu, Tiago Pinto, Selima Curci, Phuong H. Nguyen, Madeleine Gibescu, Damien Ernst, Zita A. Vale, *Sparse training theory for scalable and efficient agents,* **AAMAS** 2021

Blogs
- Michael Klear, *The Sparse Future of Deep Learning. Could this be the next quantum leap in AI research?,* https://towardsdatascience.com/the-sparse-future-of-deep-learning-bce05e8e094a, 2018
- Anhinga*, Regularization in intrinsically sparse networks,* https://github.com/anhinga/synapses/blob/master/regularization.md , 2019
- Google AI, https://ai.googleblog.com/2020/09/improving-sparse-training-with-rigl.html, 2020
- Numenta, Lucas Souza, https://numenta.com/blog/2020/10/30/case-for-sparsity-in-neural-networks-part-2-dynamic-sparsity, 2020
- Hugo Tessler, https://towardsdatascience.com/neural-network-pruning-101-af816aaea61, 2021

PIECE 1 OF THE PUZZLE
SPARSE TRAINING – OPEN-SOURCE CODE EXAMPLES

From our team

Simulating sparsity with binary masks
- https://github.com/dcmocanu/sparse-evolutionary-artificial-neural-networks
- https://github.com/Shiweiliuiiiiiii/In-Time-Over-Parameterization
- https://github.com/VITA-Group/FreeTickets
- https://github.com/VITA-Group/Random_Pruning
- https://github.com/zahraatashgahi/CTRE
- https://github.com/QiaoXiao7282/DSN

Truly sparse implementations:
- https://github.com/dcmocanu/sparse-evolutionary-artificial-neural-networks
- https://github.com/SelimaC/large-scale-sparse-neural-networks

Sparse neural networks understanding:
- https://github.com/Shiweiliuiiiiiii/Sparse_Topology_Distance

Feature selection with truly sparse implementation:
- https://github.com/zahraatashgahi/QuickSelection

For continual learning
- https://github.com/GhadaSokar/SpaceNet
- https://github.com/GhadaSokar/AFAF

For deep reinforcement learning
- https://github.com/GhadaSokar/Dynamic-Sparse-Training-for-Deep-Reinforcement-Learning
- https://github.com/bramgrooten/automatic-noise-filtering

From the community

Simulating sparsity with binary mask
- https://github.com/IntelAI/dynamic-reparameterization
- https://github.com/TimDettmers/sparse_learning
- https://github.com/google-research/rigl
- https://github.com/boone891214/MEST

FUTURE AIMS

These are difficult challenges. There is hard work ahead of us!

- ✓ for all common dense ANN architectures:
  - ➢ corresponding sparse ANNs
  - ➢ truly sparse software framework
- ❖ over 10x energy consumption and costs reduction
- ❖ over 10x reduction in CO_2 deep learning emissions
- ❖ enabling ANN training at the edge

- ✓ formal basis of sparse training theory:
  - ➢ increased performance
  - ➢ understanding
  - ➢ trust
- ❖ increased user confidence for sensitive tasks
- ❖ push AI knowledge well beyond state-of-the-art
- ❖ ANN paradigm shift

- ✓ novel sparse ANN architectures:
  - ➢ efficiency
  - ➢ decentralization
  - ➢ self-organization
  - ➢ native parallelism
- ❖ enabling realistic continual learning
- ❖ enabling ANNs to address very complex problems
- ❖ enabling ANNs with billions of neurons

THANK YOU VERY MUCH FOR YOUR ATTENTION!

Acknowledgement:
- To all my teachers, mentors, students, colleagues, collaborators, friends, family, and funding agencies!
- **And special thanks to the core team members!**

Graduated PhD students

Dr. Anil Yaman (2019)
Assistant Professor
VU Amsterdam

Dr. Shiwei Liu (2022)
Cum Laude
Postdoctoral Fellow
UT Austin

Current PhD students

Ghada Sokar
TU Eindhoven

Zahra Atashgahi
University of Twente

Qiao Xiao
TU Eindhoven

Bram Grooten
TU Eindhoven

Boqian Wu
University of Twente

?

PhD vacancy

University of Luxembourg

decebal.mocanu@uni.lu