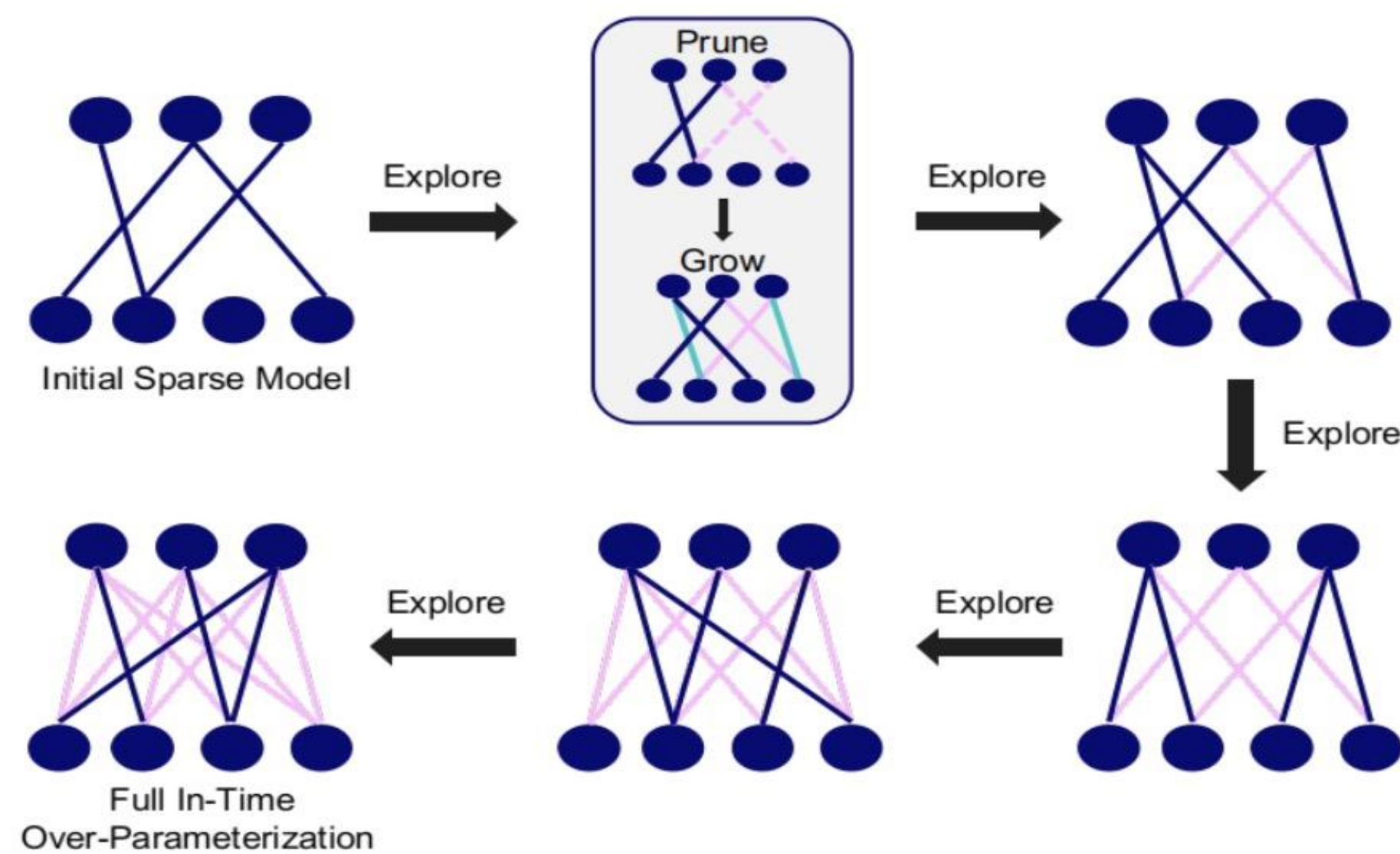# Insights into Dynamic Sparse Training: Theory Meets Practice

Boqian Wu[1,2], Maurice Van Keulen[1], Decebal Constantin Mocanu[2,3], Elena Mocanu[1]

[1]University of Twente, [2]University of Luxembourg, [3]Eindhoven University of Technology

Background

▪ **Dynamic Sparse Training (DST) [1]**

• Starting from a **sparse network** and the sparse **topology and weights** are jointly optimized.

• Applications in various domains, such as **continuous learning**, **reinforcement learning**, **feature selection**, and **network architecture design**, etc.

Research Goal

Comprehensive exploration of DST, through the three fundamental pillars of machine learning theory for supervised learning: **representation**, **optimization**, and **generalization**.

$$R(h_S) = \underbrace{\hat{R}_S(h_{S,ERM})}_{\text{representation error}} + \underbrace{\hat{R}_S(h_S) - \hat{R}_S(h_{S,ERM})}_{\text{optimization error}} + \underbrace{R(h_S) - \hat{R}_S(h_S)}_{\text{generalization error}}$$
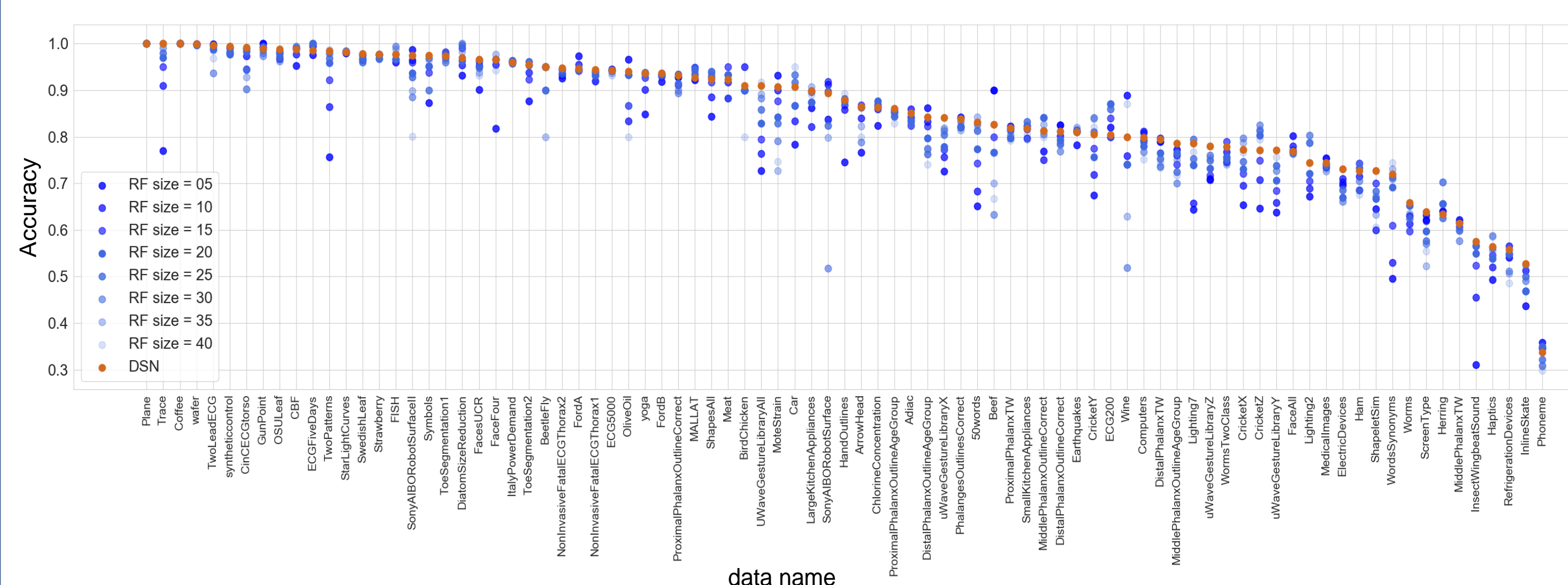
Representation

Question 1

What **specific knowledge is encoded** in sparse neural networks (SNNs) during DST?

• **Spatial domain**

  ➤ DST effectively generates a diverse Effective Receptive Field (ERF) when using large kernels [2].
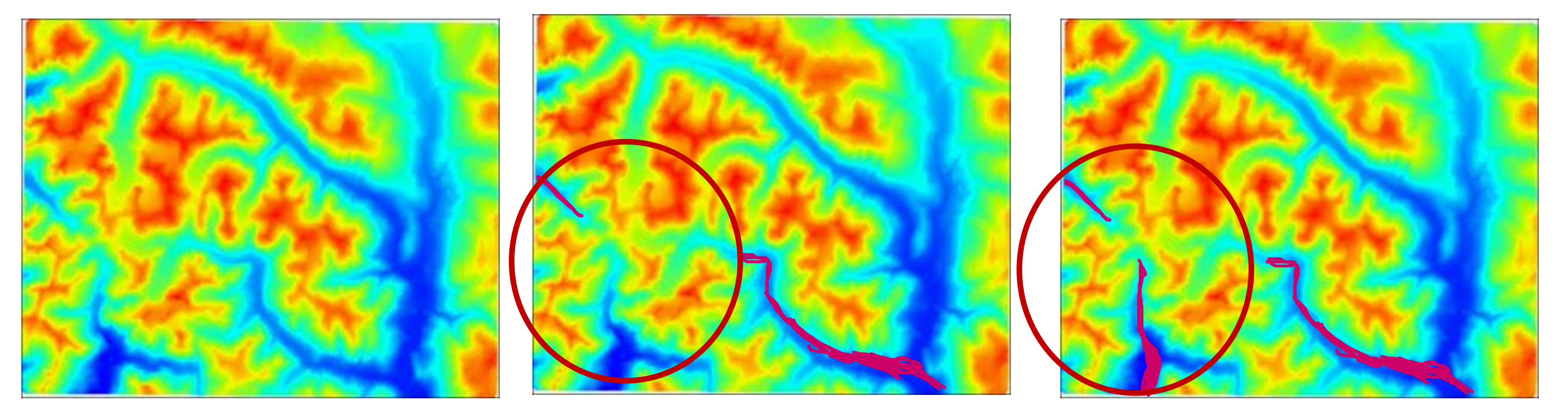

• **Spectral domain**

  ➤ DST focuses on low-frequency information, which is more perceivable to human.

Optimization

Question 2

What is the difference between the **optimization trajectories** in static sparse and dynamic sparse weight spaces?
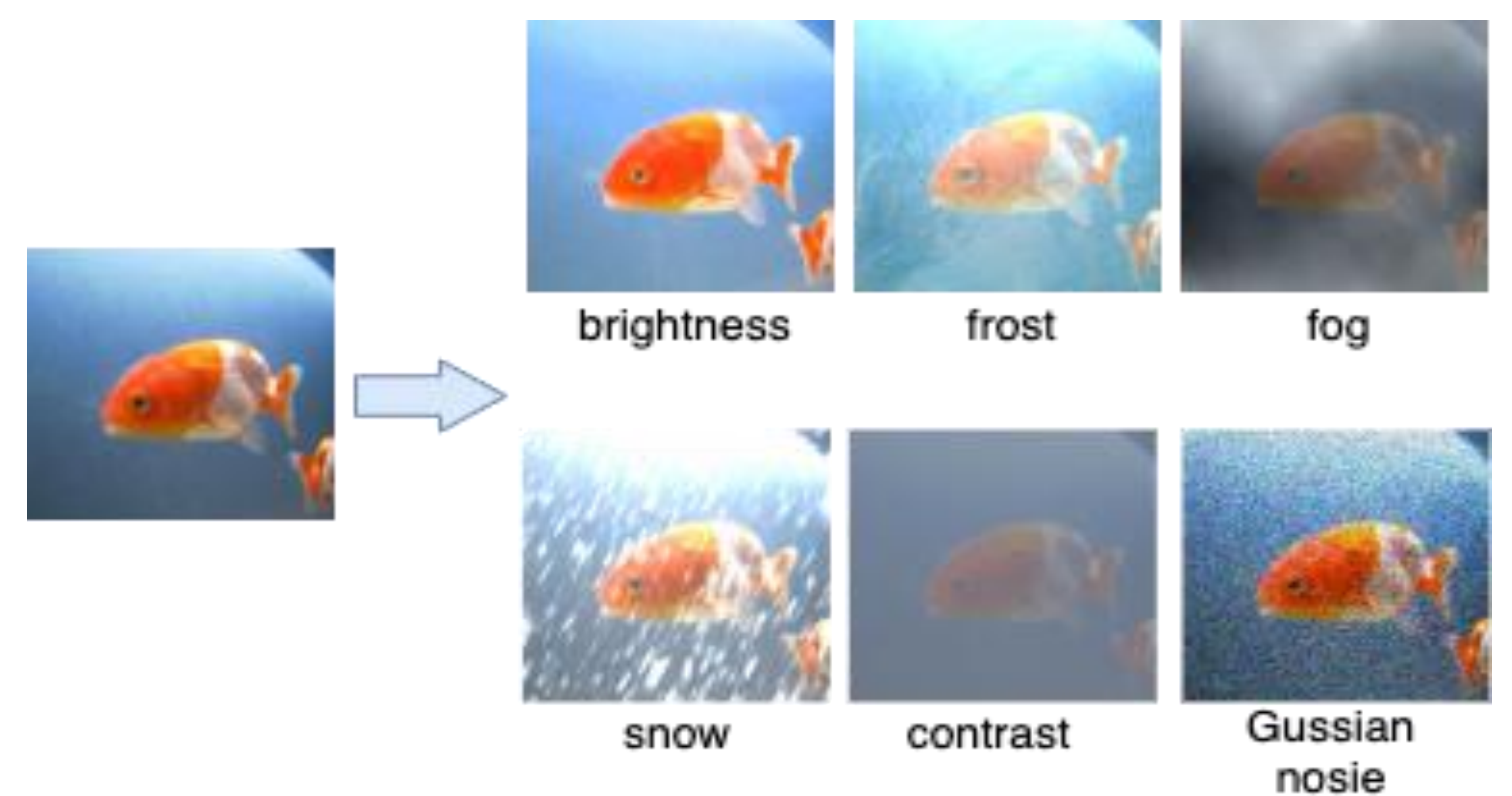
Hypothesis: DST helps to jump over a local minimum.

Genelization

Question 3

Does the reduced parameter count during DST lead to improved generalization of SNNs?

Hypothesis: DST induces the implicit regulation helps to improve generalization.

Conclusion

The ultimate goal is to harmonise the empirical achievements of DST with a **deeper theoretical comprehension**, setting a foundation to advance machine learning technologies toward more **sustainable** and **effective** solutions.

References

[1] Mocanu D C, Mocanu E, Stone P, et al. Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science[J]. Nature communications, 2018.

[2] Qiao Xiao, Boqian Wu, Yu Zhang, Shiwei Liu, Mykola Pechenizkiy, Elena Mocanu, Decebal Constantin Mocanu. Dynamic Sparse Network for Time Series Classification: Learning What to" see", NeurIPS 2022.