Backdoor Attacks on Spiking NNs and Neuromorphic Datasets

Gorka Abad [1,2] Oğuzhan Ersoy [1] Stjepan Picek [1] Víctor Julio Ramírez-Durán [2] Aitor Urbieta [2]

ACM CCS 2022

[1] Radboud University

[2] Ikerlan Technology Research Centre

Table of Contents

Introduction

- What is a SNN?
- What is Neuromorphic data?

Backdoor Attacks

- Backdoor Attacks in DL
- How to Backdoor SNN
- Static Backdoor Attack
- Moving Backdoor Attack

Results

Future Work

Table of Contents

Introduction

- What is a SNN?
- What is Neuromorphic data?

Backdoor Attacks

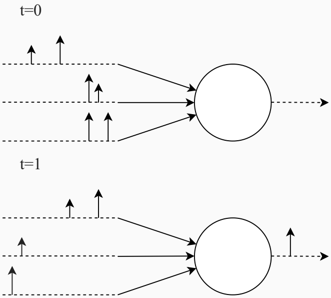
- Backdoor Attacks in DL
- How to Backdoor SNN
- Static Backdoor Attack
- Moving Backdoor Attack

Results

Future Work

Introduction: What is a SNN?

- Energy efficient NN [1]
- Instead of Neurons, SNN use *Spiking Neurons*
- Spiking neurons are excited up to a threshold
- When the threshold is reached they fire



Introduction: What is Neuromorphic data?

- Time encoded data
- Usually captured by a DVS camera
- Captures the changes in luminosity, expressed as *polarities*

Right hand wave [2]

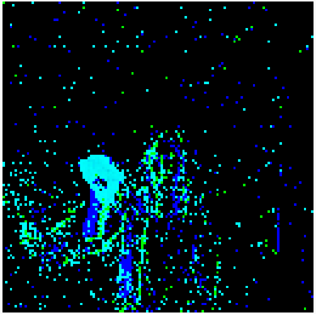


Table of Contents

Introduction

- What is a SNN?
- What is Neuromorphic data?

Backdoor Attacks

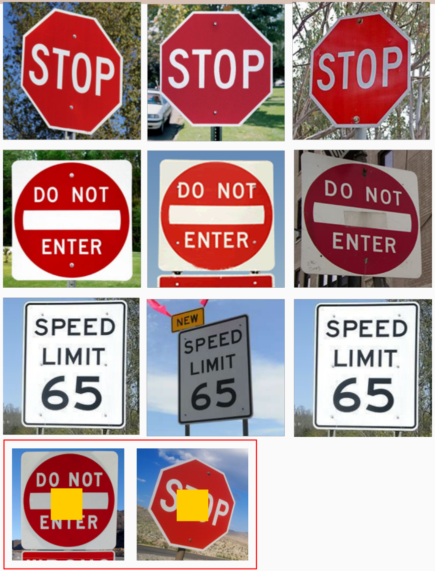
- Backdoor Attacks in DL
- How to Backdoor SNN
- Static Backdoor Attack
- Moving Backdoor Attack

Results

Future Work

Backdoor Attacks: Backdoor Attacks in DL

- What happens with untested samples?
- We can create them adding a *trigger* [3]
- Trigger: 🟨
- Label: “Speed Limit”

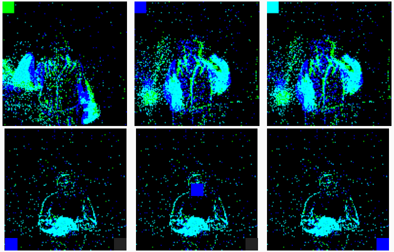


Backdoor Attacks: How to Backdoor SNN

- Samples are not images
- The pixel space is reduced to 2 bits, from (usually) 255
- We change the pixel polarities to inject different triggers
- Samples are divided in frames, i.e., movement
- 2 possible approaches
  - *Static triggers*
  - *Moving triggers*

Backdoor Attacks: Static Backdoor Attack

- Triggers are in the same place in all the frames, i.e., static
- Advantages
  - Great backdoor performance
  - Easy to implement
- Disadvantages
  - Easy to detect. They are static when neuromophic data suggest movement



Backdoor Attacks: Moving Backdoor Attack

- Triggers move from between frames
- Advantages
  - Stealthier than static triggers
  - The trigger can be placed anywhere
- Disadvantages
  - Slighter lower performance than static triggers

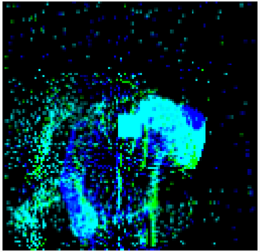


Table of Contents

- Introduction
  - What is a SNN?
  - What is Neuromorphic data?
- Backdoor Attacks
  - Backdoor Attacks in DL
  - How to Backdoor SNN
  - Static Backdoor Attack
  - Moving Backdoor Attack
- Results
- Future Work

Results: Moving Backdoor Attack

Table: Result of our attack under different settings.

Dataset	ϵ	k	s	p	Static/Moving	Epochs	Main task accuracy	Backdoor accuracy
DVS Gesture	-	-	-	-	-	65	76%	-
DVS Gesture	0.01	top-left	0.1	2	Static	65	76%	1%
DVS Gesture	0.1	middle	0.1	0	Static	65	74%	99%
DVS Gesture	**0.1**	**bottom-right**	**0.1**	**1**	**Static**	**65**	**76%**	**100%**
DVS Gesture	**0.01**	**top-left**	**0.3**	**0**	**Static**	**65**	**76%**	**99%**
DVS Gesture	**0.1**	**bottom-right**	**0.1**	**1**	**Moving**	**65**	**76%**	**99%**
N-MNIST	-	-	-	-	-	20	99%	-
N-MNIST	**0.001**	**bottom-right**	**0.1**	**0**	**Static**	**20**	**99%**	**98%**
N-MNIST	0.001	middle	0.1	1	Static	20	97%	1%
N-MNIST	0.01	top-left	0.1	0	Static	20	98%	100%
N-MNIST	**0.001**	**middle**	**0.1**	**0**	**Moving**	**20**	**98%**	**93%**

Table of Contents

Introduction

- What is a SNN?
- What is Neuromorphic data?

Backdoor Attacks

- Backdoor Attacks in DL
- How to Backdoor SNN
- Static Backdoor Attack
- Moving Backdoor Attack

Results

Future Work

Future Work: Moving Backdoor Attack

- Defences
- Dynamic triggers
- Measuring and improving stealthiness

Thanks for your attention, any questions?

References

[1] Wei Fang, Zhaofei Yu, Yanqi Chen, et al. "Incorporating learnable membrane time constant to enhance learning of spiking neural networks". In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021, pp. 2661–2671.

[2] Arnon Amir, Brian Taba, David Berg, et al. "A low power, fully event-based gesture recognition system". In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp. 7243–7252.

[3] Tianyu Gu, Kang Liu, Brendan Dolan-Gavitt, et al. "Badnets: Evaluating backdooring attacks on deep neural networks". In: IEEE Access 7 (2019), pp. 47230–47244.