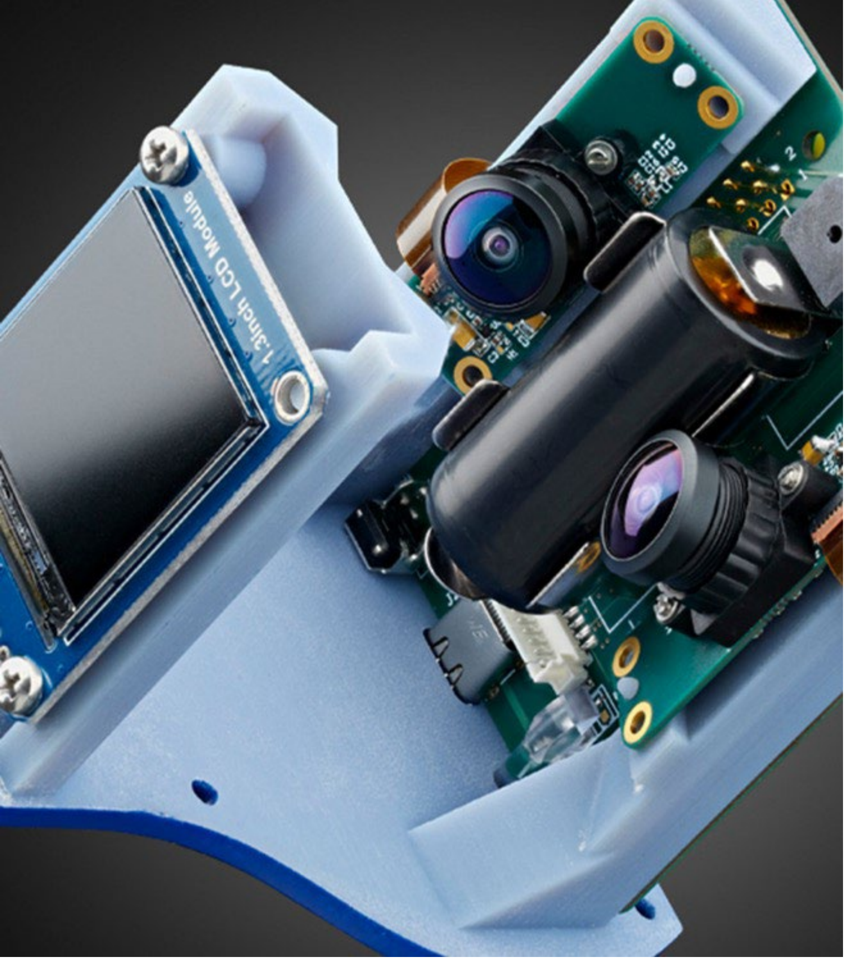
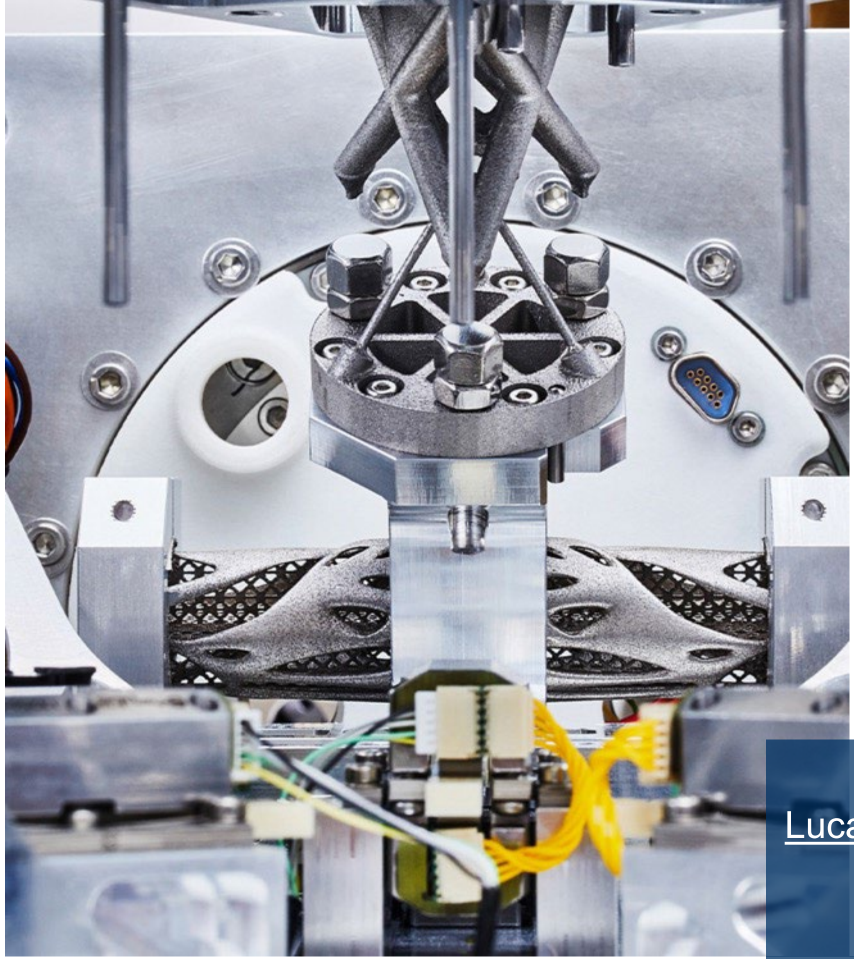




Lucas David Meier, Damian Vizár, Felipe Valencia,
Cristian-Alexandru Botocan

03/04/2025

TAKING AI-BASED SIDE-CHANNEL ATTACKS TO A NEW DIMENSION

csem

HOW TO RUN SIDE-CHANNEL ATTACKS

- Attack AES256 key byte-per-byte
  - Specific intermediate value

- [Opt.] Apply a labelling function on top
  - Hamming Weight (HW)
  - Hamming Distance

→ HW can be more correlated to the actual power leakage, but is also less informative:

Algorithm 1 AES Encryption

function AESENCRYPTION(ptx, K)
  $K_0, K_1, \ldots, K_{N_r}$ = KeyExpansion(K, N_r)
  AddRoundKey(ptx, K_0)
  for $r = 1, 2, 3, \ldots, N_r - 1$ **do**
    SubBytes(ptx)
    ShiftRows(ptx)
    MixColumns(ptx)
    AddRoundKey(ptx, K_r)
  SubBytes(ptx)
  ShiftRows(ptx)
  AddRoundKey(ptx, K_{N_r})

HW value	0	1	2	3	4	5	6	7	8
Occurrences	1	8	28	56	70	56	28	8	1

SIDE CHANNEL ATTACKS (SCA) & AI

- A lot of publications in the past 15 years
- Improvements are mainly about:
  - Re-using generic AI techniques and applying them to SCA
    - Optimizers [1]
    - Vizualisation (in unprofiled attacks) [2]
    - Learning Rates [3, 4]
  - Model selection and fine-tuning [5, 6, 7, 8]
  - Pre-processing operations (“Make Some Noise”, “Auto-encoders”, “Mean”, …) [9, 10]
- All use the same batch of public datasets to compare against each other:
  - ASCAD variants, AES_HD, AES_RD, DPAContestV4, CHES CTF 2023 (SMAesH), …

[1] Perin, G., Picek, S.: On the Influence of Optimizers in Deep Learning-based Side-channel Analysis. In: Cryptology ePrint Archive, Paper 2020/977 (2020)
[2] Timon, B.: Non-Profiled Deep Learning-based Side-Channel attacks with Sensitivity Analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems 2019(2), 107–131 (Feb 2019). https://doi.org/10.13154/tches.v2019.i2.107-131, https://tches.iacr.org/index.php/TCHES/article/view/7387
[3] Smith, L.N.: Cyclical learning rates for training neural networks. In: 2017 IEEE winter conference on applications of computer vision (WACV). pp. 464–472. IEEE (2017)
[4] Masure, L., Dumas, C., Prouff, E.: Gradient Visualization for General Characterization in Profiling Attacks. In: Constructive Side-Channel Analysis and Secure Design, pp. 145–167. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-16350-19, https://doi.org/10.1007/978-3-030-16350-1_9
[5] Wu, L., Perin, G., Picek, S.: I Choose You: Automated Hyperparameter Tuning for Deep Learning-based Side-channel Analysis. Cryptology ePrint Archive, Report 2020/1293 (2020), https://ia.cr/2020/1293
[6] Rijsdijk, J., Wu, L., Perin, G., Picek, S.: Reinforcement learning for hyperparameter tuning in deep learning-based side-channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems pp. 677–707 (2021)
[7] Wouters, L., Arribas, V., Gierlichs, B., Preneel, B.: Revisiting a methodology for efficient CNN architectures in profiling attacks. IACR Transactions on Cryptographic Hardware and Embedded Systems pp. 147–168 (2020)
[8] Zaid, G., Bossuet, L., Habrard, A., Venelli, A.: Methodology for Efficient CNN Architectures in Profiling Attacks (Nov 2019). https://doi.org/10.13154/tches.v2020.i1.1-36, https://tches.iacr.org/index.php/TCHES/article/view/8391
[9] Wu, L., Picek, S.: Remove some noise: On pre-processing of side-channel measurements with autoencoders. IACR Transactions on Cryptographic Hardware and Embedded Systems pp. 389–415 (2020)
[10] Wu, L., & Picek, S. (2020). Remove Some Noise: On Pre-processing of Side-channel Measurements with Autoencoders. IACR Cryptol. ePrint Arch., 2019, 1474.

SIDE-CHANNEL ATTACKS (SCA) & CLASS IMBALANCE

- Picek, S. et al. [1] proposed
  - SMOTE as a best-working solution to combat imbalanced datasets in the SCA context
    - SMOTE generates artificial samples (over-sampling) of rare classes in the profiling set to even-out all classes
  - Not to use labelling (i.e. Identity labelling) for best attack performance

- Since then, only *few* papers proposed a comparison using HW

    - What if… there was more to HW ?

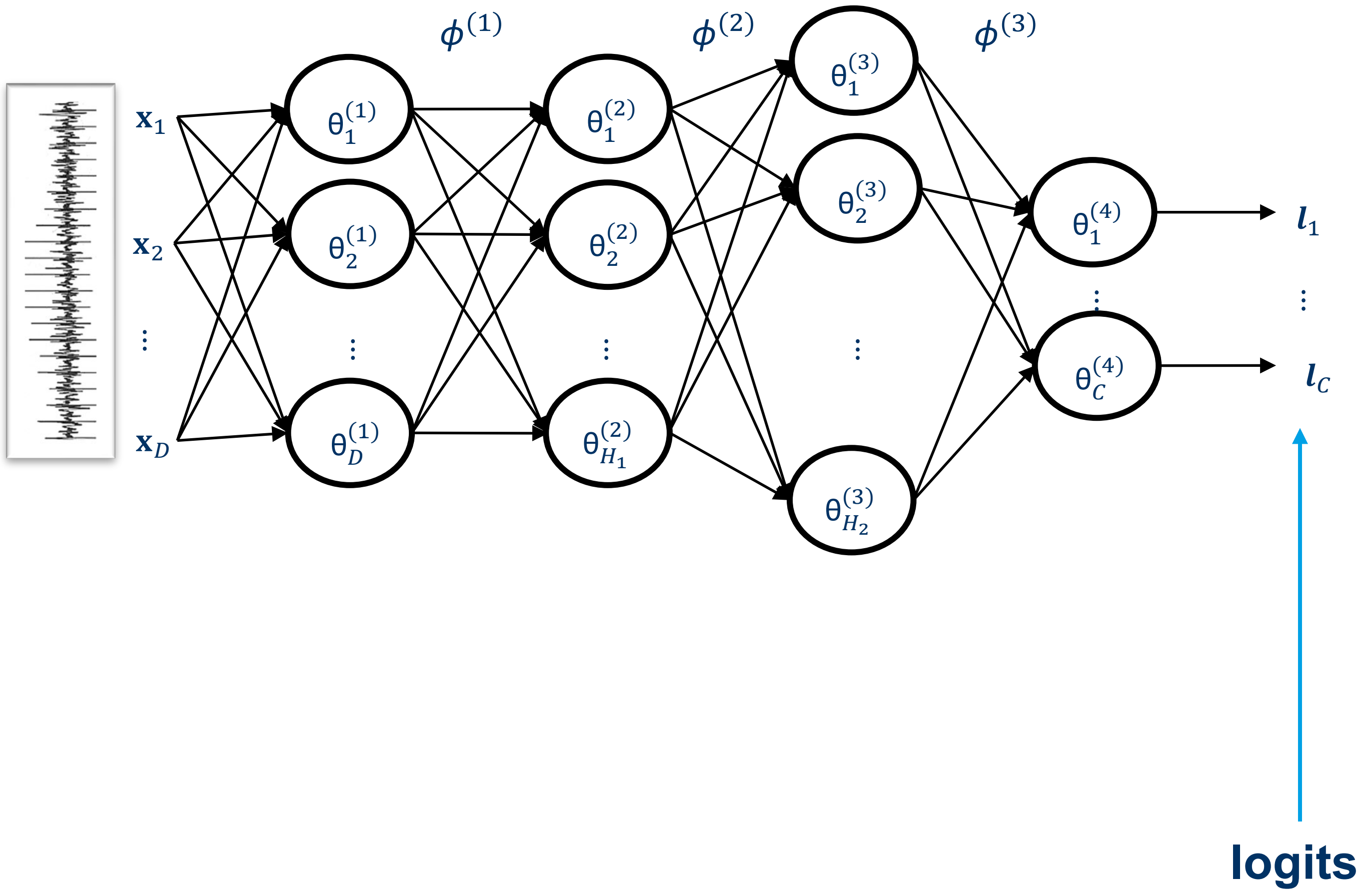
[1] Picek, S., Heuser, A., Jovic, A., Bhasin, S., Regazzoni, F.: The Curse of Class Imbalance and Conflicting Metrics with Machine Learning for Side-channel Evaluations. IACR Transactions on Cryptographic Hardware and Embedded Systems 2019(1), 1–29 (Aug 2019). https://doi.org/10.13154/tches.v2019.i1.209-237, https://hal.inria.fr/hal-01935318

TAKING AI-BASED SIDE-CHANNEL ATTACK TO A NEW DIMENSION

DEEP LEARNING BASICS

- A Multi-Layer Perceptron

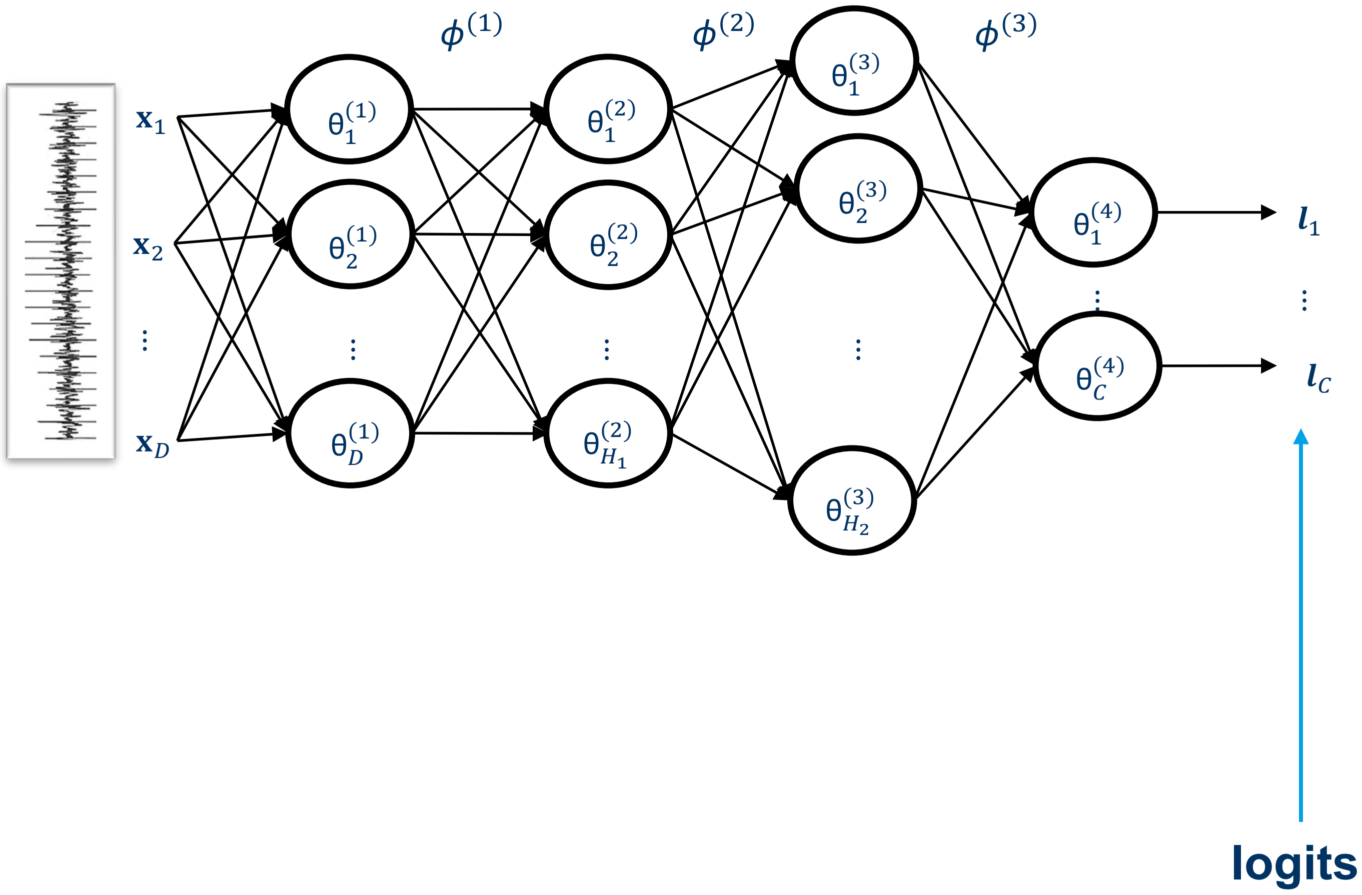
$\theta_i^{(j)}$: **i-th parameter of layer j**



DEEP LEARNING BASICS

- A Multi-Layer Perceptron

$\theta_i^{(j)}$: **i-th parameter of layer j**



Each logit corresponds to one class

SOFTMAX FUNCTIONS

- The **softmax** function normalizes the logits, each prediction will sum to 1
- $Softmax(\mathrm{L}, t, c) = \frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$
- Each power trace is mapped to a probability density function over the different classes
- In the implementation, the softmax function is called once over a 2D matrix of logits L of size (Batch-size / # output-classes)

csem

SOFTMAX FUNCTIONS

- The **softmax** function normalizes the logits, each prediction will sum to 1
- $Softmax(\mathrm{L}, t, c) = \dfrac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$
- Each power trace is mapped to a probability density function over the different classes
- In the implementation, the softmax function is called once over a 2D matrix of logits L of size (Batch-size / # output-classes)
- Our work shows that transposing this input matrix confers promising properties for a SCA
  - We dubbed this variant “Dimension 0”

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0			
t_1			
t_2			
t_3			

DIM = 1	c_0	c_1	c_2
t_0			
t_1			
t_2			
t_3			

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0			
t_1			
t_2			
t_3			

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1			
t_2			
t_3			

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0			
t_1			
t_2			
t_3			

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2			
t_3			

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0			
t_1			
t_2			
t_3			

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2	0 | 0%	59 | 100%	1 | 0%
t_3			

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0			
t_1			
t_2			
t_3			

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2	0 | 0%	59 | 100%	1 | 0%
t_3	5 | 0%	55 | 100%	0 | 0%

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0	[redacted]	[redacted]	[redacted]
t_1	[redacted]	[redacted]	[redacted]
t_2	[redacted]	[redacted]	[redacted]
t_3	[redacted]	[redacted]	[redacted]

<table>
<tr><th>DIM = 1</th><th>c_0</th><th>c_1</th><th>c_2</th></tr>
<tr><td>t_0</td><td>0 | 0%</td><td>44 | 100%</td><td>7 | 0%</td></tr>
<tr><td>t_1</td><td>0 | 0%</td><td>36 | 100%</td><td>2 | 0%</td></tr>
<tr><td>t_2</td><td>0 | 0%</td><td>59 | 100%</td><td>1 | 0%</td></tr>
<tr><td>t_3</td><td>5 | 0%</td><td>55 | 100%</td><td>0 | 0%</td></tr>
</table>

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0	2 | 0%		
t_1	3 | 0%		
t_2	0 | 0%		
t_3	9 |100%		

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2	0 | 0%	59 | 100%	1 | 0%
t_3	5 | 0%	55 | 100%	0 | 0%

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0	2 | 0%	7 | 0%	[redacted]
t_1	3 | 0%	0 | 0%	[redacted]
t_2	0 | 0%	20 | 50%	[redacted]
t_3	9 |100%	20 | 50%	[redacted]

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2	0 | 0%	59 | 100%	1 | 0%
t_3	5 | 0%	55 | 100%	0 | 0%

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0	2 | 0%	7 | 0%	8 |97%
t_1	3 | 0%	0 | 0%	4 | 2%
t_2	0 | 0%	20 | 50%	0 | 0 %
t_3	9 |100%	20 | 50%	3 | 1 %

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2	0 | 0%	59 | 100%	1 | 0%
t_3	5 | 0%	55 | 100%	0 | 0%

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0	2 | 0%	7 | 0%	8 |97%
t_1	3 | 0%	0 | 0%	4 | 2%
t_2	0 | 0%	20 | 50%	0 | 0 %
t_3	9 |100%	20 | 50%	3 | 1 %

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2	0 | 0%	59 | 100%	1 | 0%
t_3	5 | 0%	55 | 100%	0 | 0%

$$\frac{e^{L_{t,c}}}{\sum_{j=0}^{n} e^{L_{t,j}}}$$

csem

DIMENSION 0 - EXAMPLE

Original	c_0	c_1	c_2
t_0	3	47	10
t_1	4	40	6
t_2	1	60	2
t_3	10	60	5

DIM = 0	c_0	c_1	c_2
t_0	2 | 0%	7 | 0%	8 |97%
t_1	3 | 0%	0 | 0%	4 | 2%
t_2	0 | 0%	20 | 50%	0 | 0 %
t_3	9 |100%	20 | 50%	3 | 1 %

DIM = 1	c_0	c_1	c_2
t_0	0 | 0%	44 | 100%	7 | 0%
t_1	0 | 0%	36 | 100%	2 | 0%
t_2	0 | 0%	59 | 100%	1 | 0%
t_3	5 | 0%	55 | 100%	0 | 0%

- **Proposition 1**
  - Per-class processing
  - Increase a class score (logit-value) for an input ←→ decreasing from another input

csem

DIMENSION 0 - INSIGHTS

- **Corollary 1**
  1. Class score imbalance is not preserved

Original	c_0	c_1	c_2
t_1	4	40	6

DIM = 0	c_0	c_1	c_2
t_1	3 | 0%	0 | 0%	4 | 2%

  2. Elect best input representatives for every class
     → Usually, it's the opposite
  3. Better consideration of rare classes

csem

DIMENSION 0’S INCREASED CONSIDERATION TO RARE CLASSES

- The sum of the class-scores for one input trace does not sum to 1 anymore
- However, the global key-ranking algorithm did not change
  - Consequence:

DIM = 0	c_0	c_1	c_2
t_0	2 | 0%	7 | 0%	8 |97%
t_1	3 | 0%	0 | 0%	4 | 2%
t_2	0 | 0%	20 | 50%	0 | 0 %
t_3	9 |100%	20 | 50%	3 | 1 %

This trace accounts only for 2% on C2

This trace accounts for 100% on C0, 50% on C1 and 1% on C2

csem

DIMENSION 0'S INCREASED CONSIDERATION TO RARE CLASSES

- The sum of the class-scores for one input trace does not sum to 1 anymore
- However, the global key-ranking algorithm did not change
  - Consequence:

DIM = 0	c_0	c_1	c_2
t_0	2 | 0%	7 | 0%	8 |97%
t_1	3 | 0%	0 | 0%	4 | 2%
t_2	0 | 0%	20 | 50%	0 | 0 %
t_3	9 |100%	20 | 50%	3 | 1 %

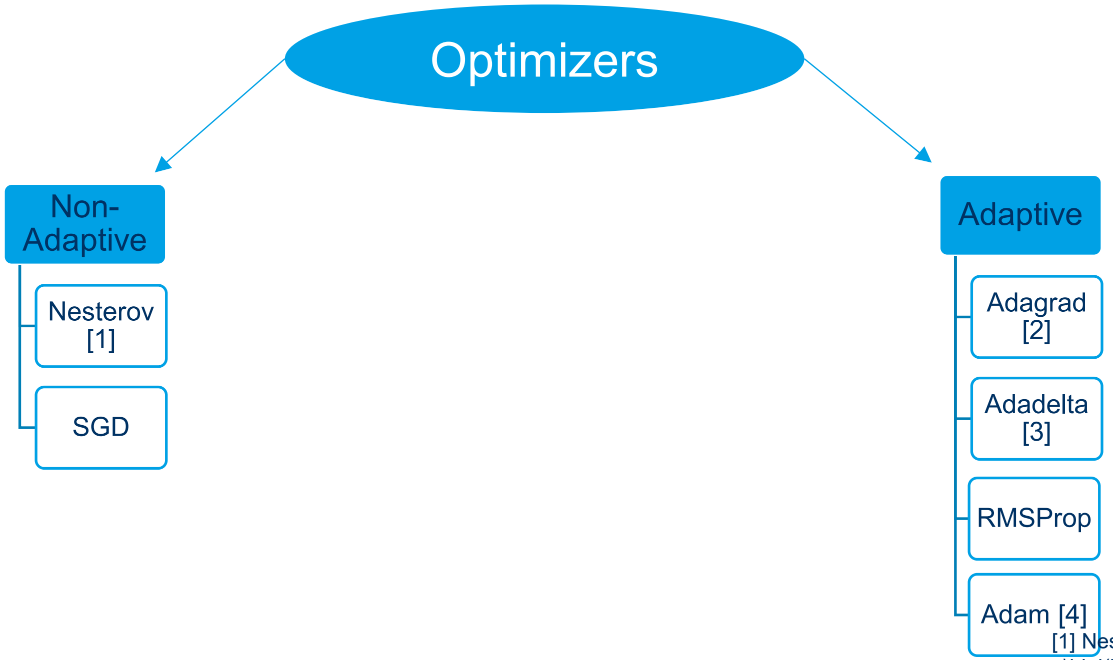
This trace accounts only for 2% on C2

This trace accounts for 100% on C0, 50% on C1 and 1% on C2

→ Input traces with *easily classifiable classes* tend to have more weight

MAXIMIZING DIMENSION 0 PERFORMANCE

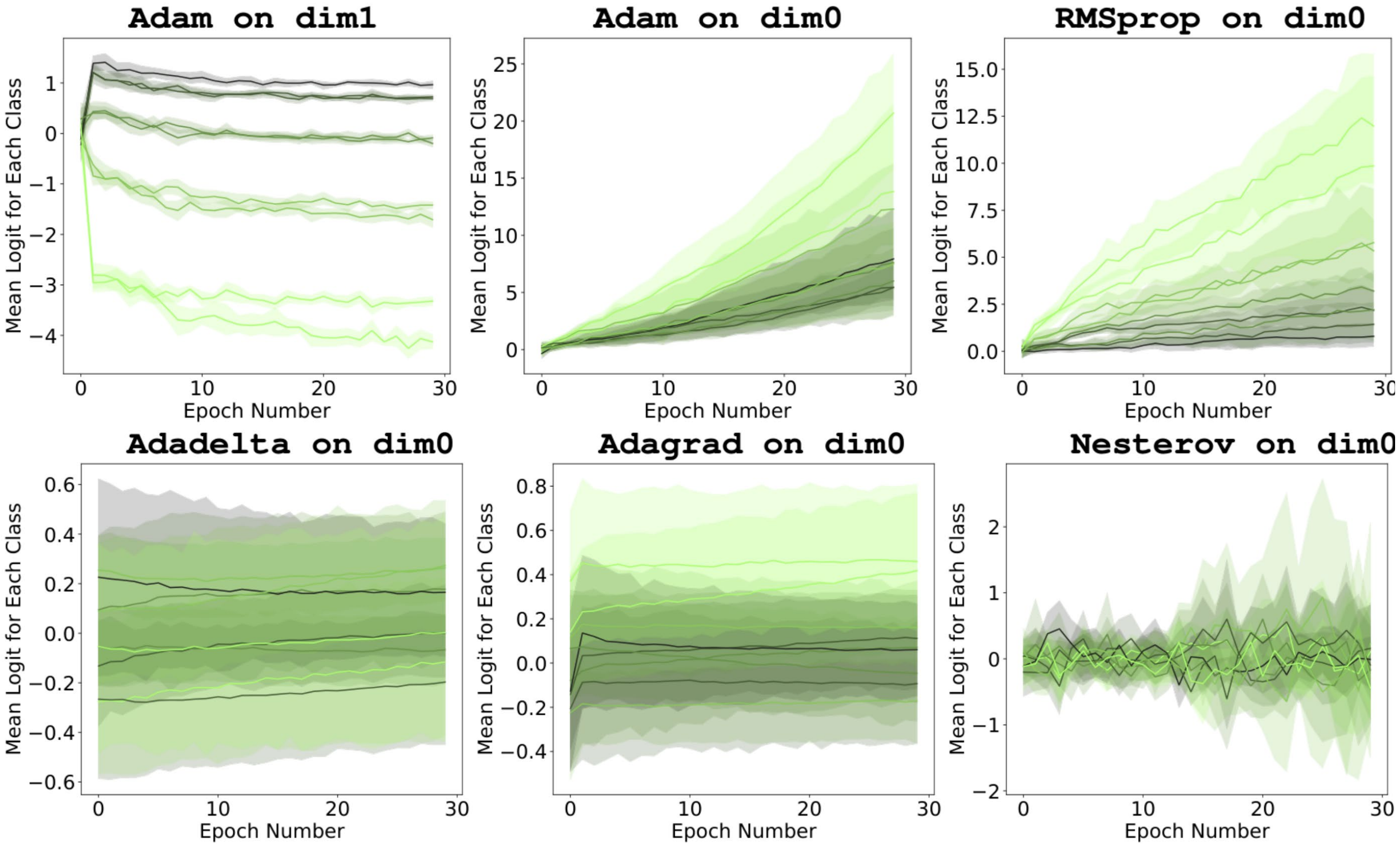
- We assess Dimension 0 performance with different optimizers
  - Adaptive optimizers have, for each batch, an adaptive learning rate for each parameter.



[1] Nesterov, Y.E.: A method of solving a convex programming problem with convergence rate o\bigl(k^2\bigr). In: Doklady Akademii Nauk. vol. 269, pp. 543–547. Russian Academy of Sciences (1983)
[2] Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. Journal of machine learning research 12(7) (2011)
[3] Zeiler, M.D.: Adadelta: An adaptive learning rate method (2012)
[4] Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization (2014)

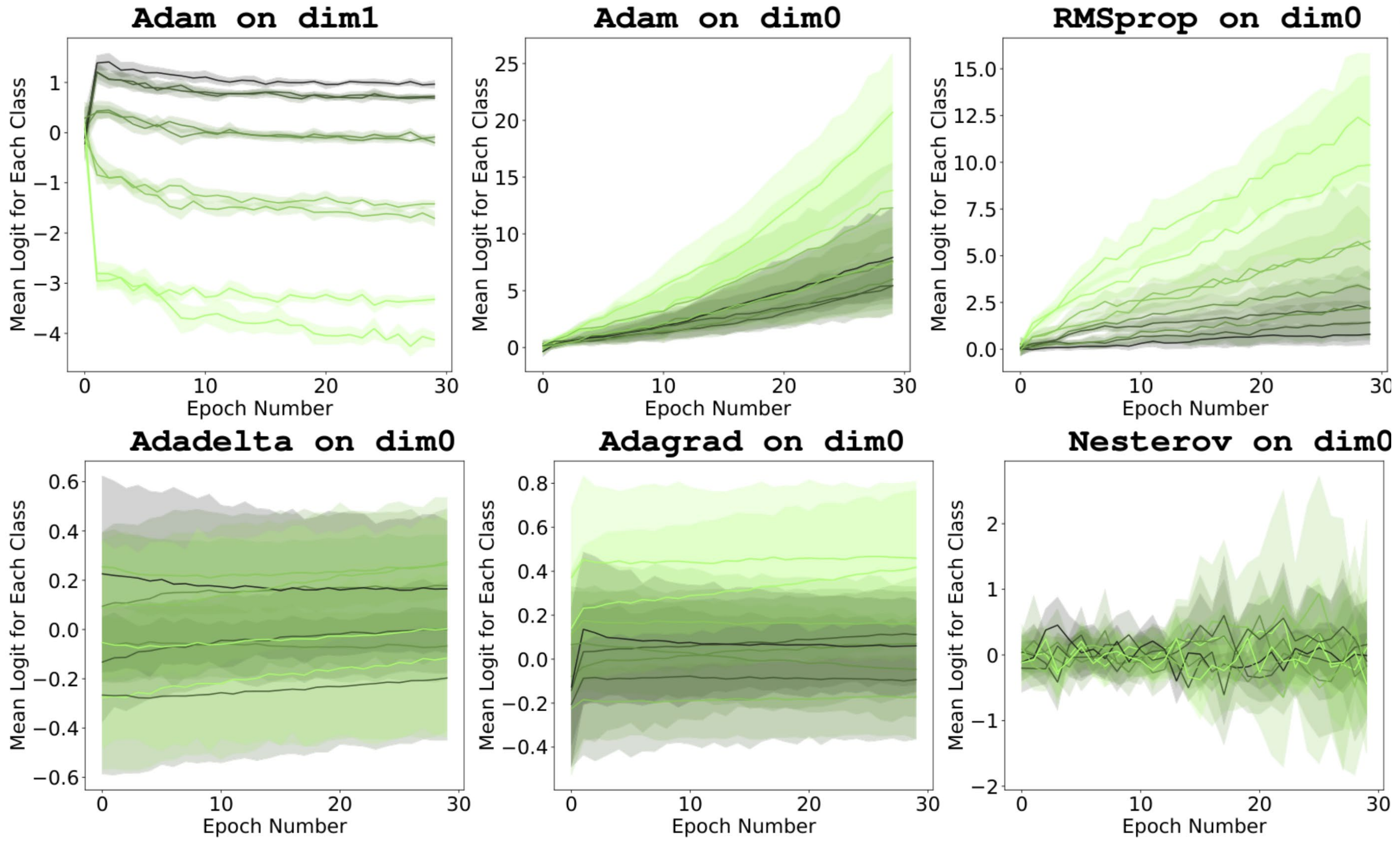
csem

MAXIMIZING DIMENSION 0 PERFORMANCE



Mean logit value for each class of first batch at each epoch.
Light green ←→ rare classes.
Dark green ←→ common classes.
Model: CNN_exp. Dataset: AES_nRF.

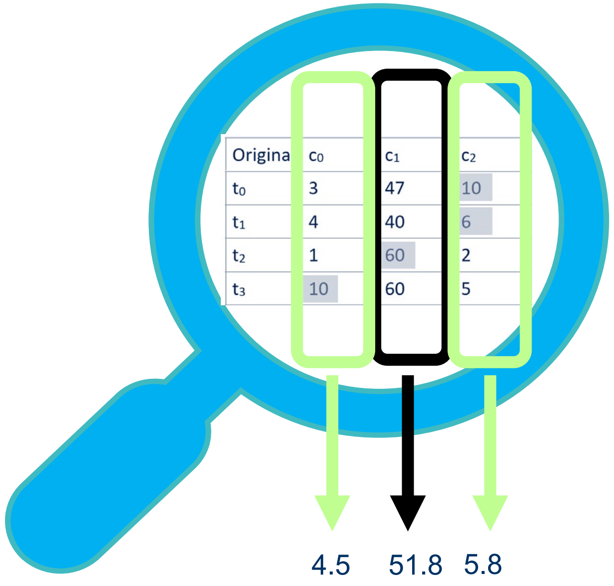
MAXIMIZING DIMENSION 0 PERFORMANCE

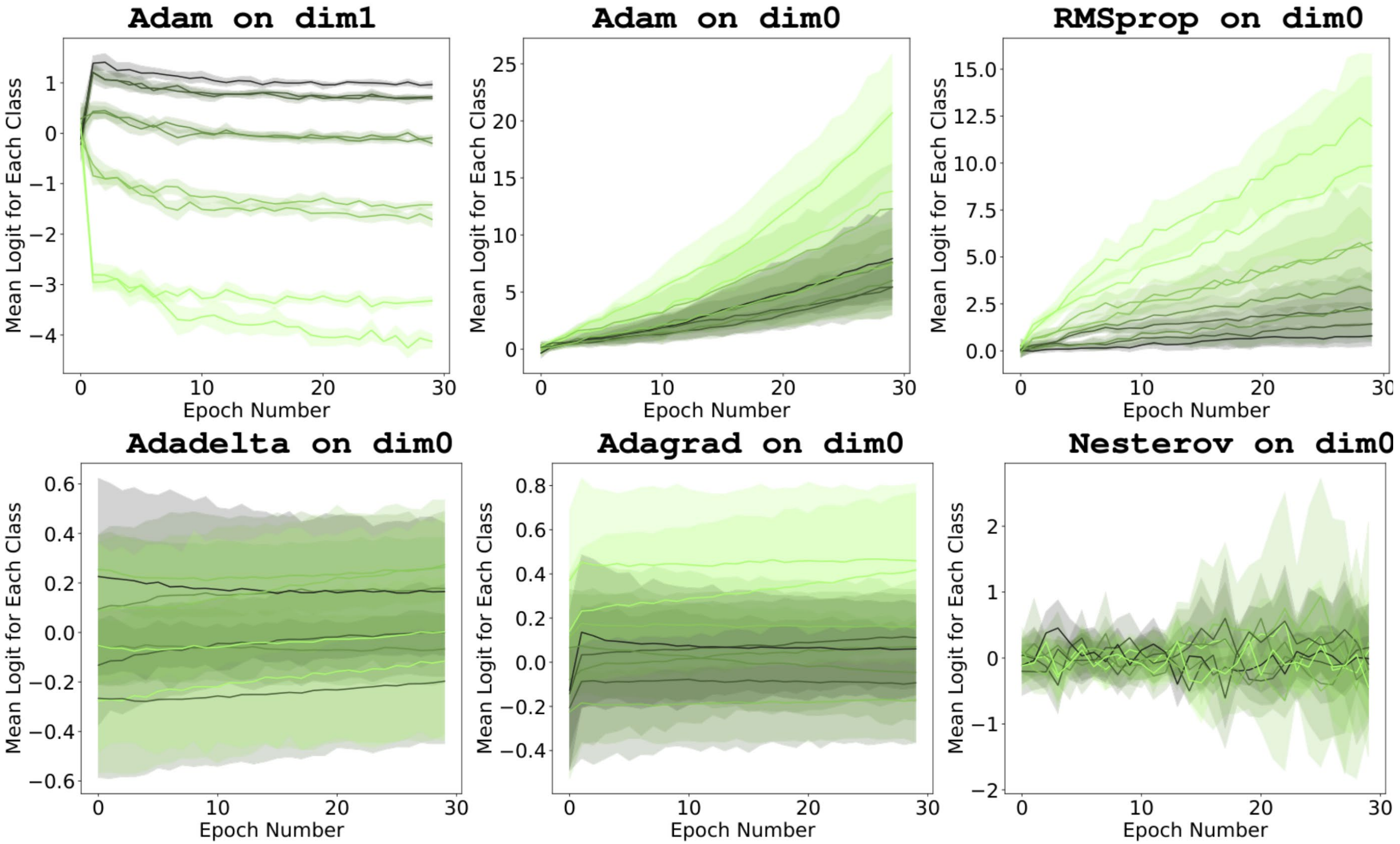


Mean logit value for each class of first batch at each epoch.
Light green ←→ rare classes.
Dark green ←→ common classes.
Model: CNN_exp. Dataset: AES_nRF.

csem

MAXIMIZING DIMENSION 0 PERFORMANCE





Mean logit value for each class of first batch at each epoch.
Light green ←→ rare classes.
Dark green ←→ common classes.
Model: CNN_exp. Dataset: AES_nRF.

COMBINING BENEFITS

Dimension 0

- Train each class separately

Adaptive Optimizer

- Train each weight separately

Benefits

- Rare classes are better off
- Traces with targeted labels have more weight in ranking

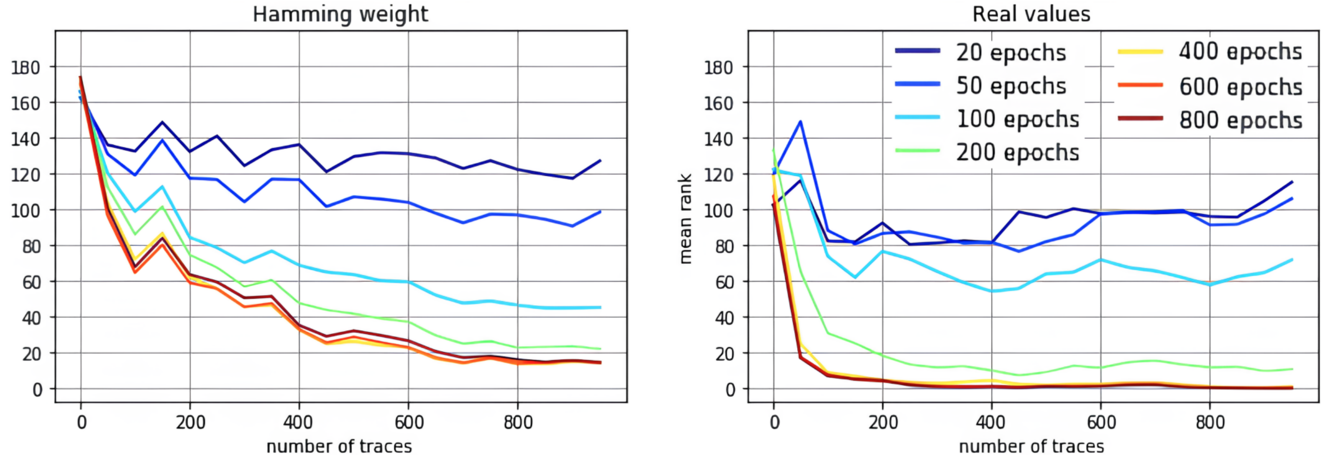
csem

EXPERIMENTS

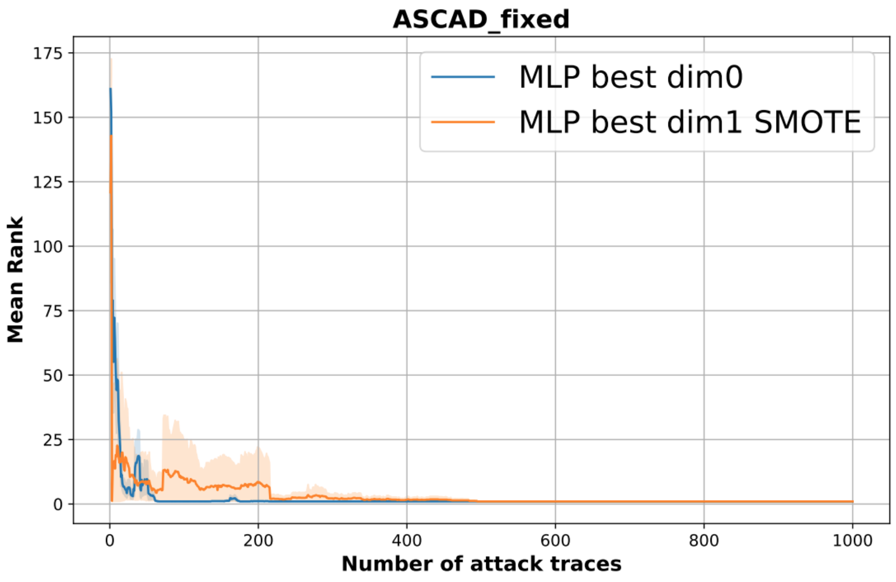
ASCAD: SIMPLE, FIXED KEY

Technique	Approx. number of traces	Comment
Dim 1 with ID	300	
Dim 1 with HW	∞	Not feasible
Dim 1 with HW + SMOTE	450	Noisy
Dim 0 with HW	200	100 may be enough

Attacking ASCAD_fixed with the MLP_best model and batch size 100



(a) ASCAD_fixed, MLP_best results [1] with HW (left) and ID (right)



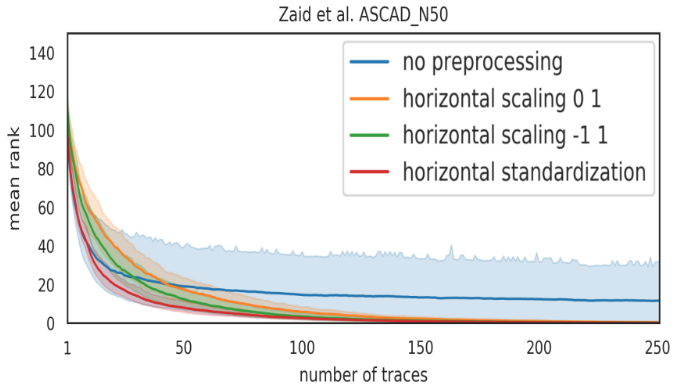
(b) ASCAD_fixed, MLP_best results dim0 and dim1 with SMOTE with HW labelling

[1] Benadjila, R., Prouff, E., Strullu, R., Cagli, E., Dumas, C.: Deep learning for side-channel analysis and introduction to ASCAD database. Journal of Cryptographic Engineering 10(2), 163–188 (Nov 2019). https://doi.org/10.1007/s13389-019-00220-8, https://doi.org/10.1007/s13389-019-00220-8

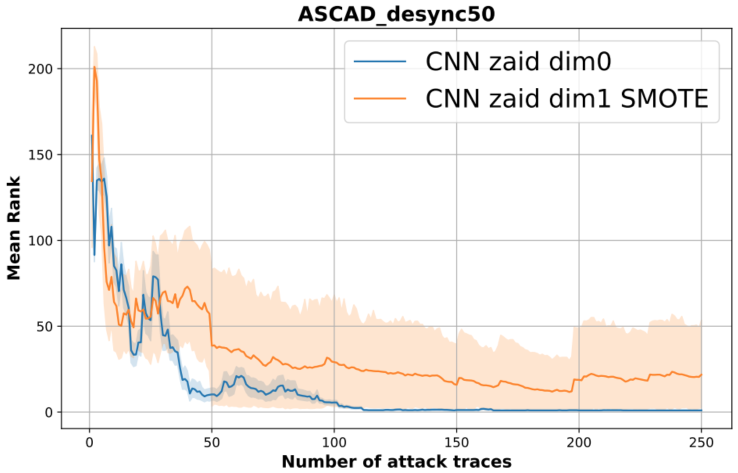
ASCAD – DESYNC LEVEL 50

Technique	Approx. number of traces	Comment
Dim 1 with ID	150	
Dim 1 with HW	N/A	
Dim 1 with HW + SMOTE	∞	Not feasible
Dim 0 with HW	<175	150 may be enough

Attacking ASCAD_desync50 with the CNN_zaid model and batch size 50



(a) ASCAD_desync50, CNN_zaid results with ID labelling [1]



(b) ASCAD_desync50, CNN_zaid results with HW labelling

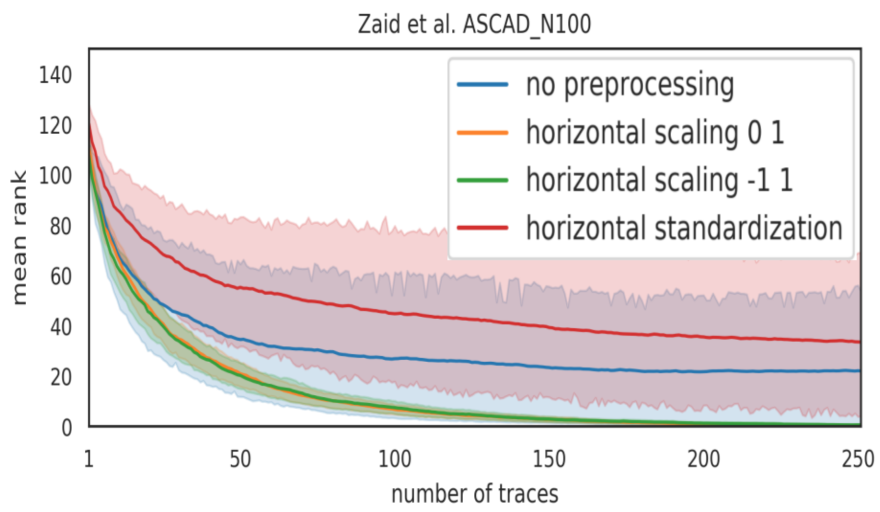
[1] Wouters, L., Arribas, V., Gierlichs, B., Preneel, B.: Revisiting a methodology for efficient CNN architectures in profiling attacks. IACR Transactions on Cryptographic Hardware and Embedded Systems pp. 147–168 (2020)

csem

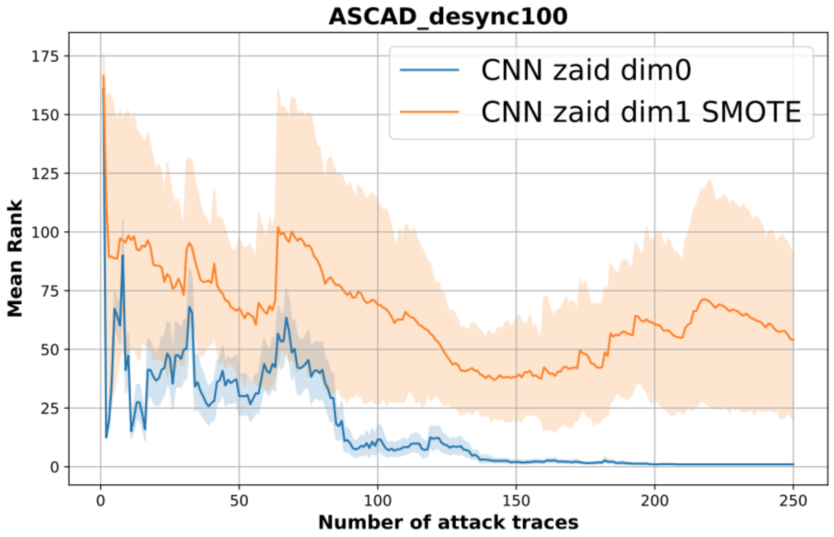
ASCAD – DESYNC LEVEL 100

Technique	Approx. number of traces	Comment
Dim 1 with ID	200	
Dim 1 with HW	N/A	
Dim 1 with HW + SMOTE	∞	Not feasible
Dim 0 with HW	200	

Attacking ASCAD_desync100 with the CNN_zaid model, batch size 50



(a) ASCAD_desync100, CNN_zaid results with ID labelling [1]

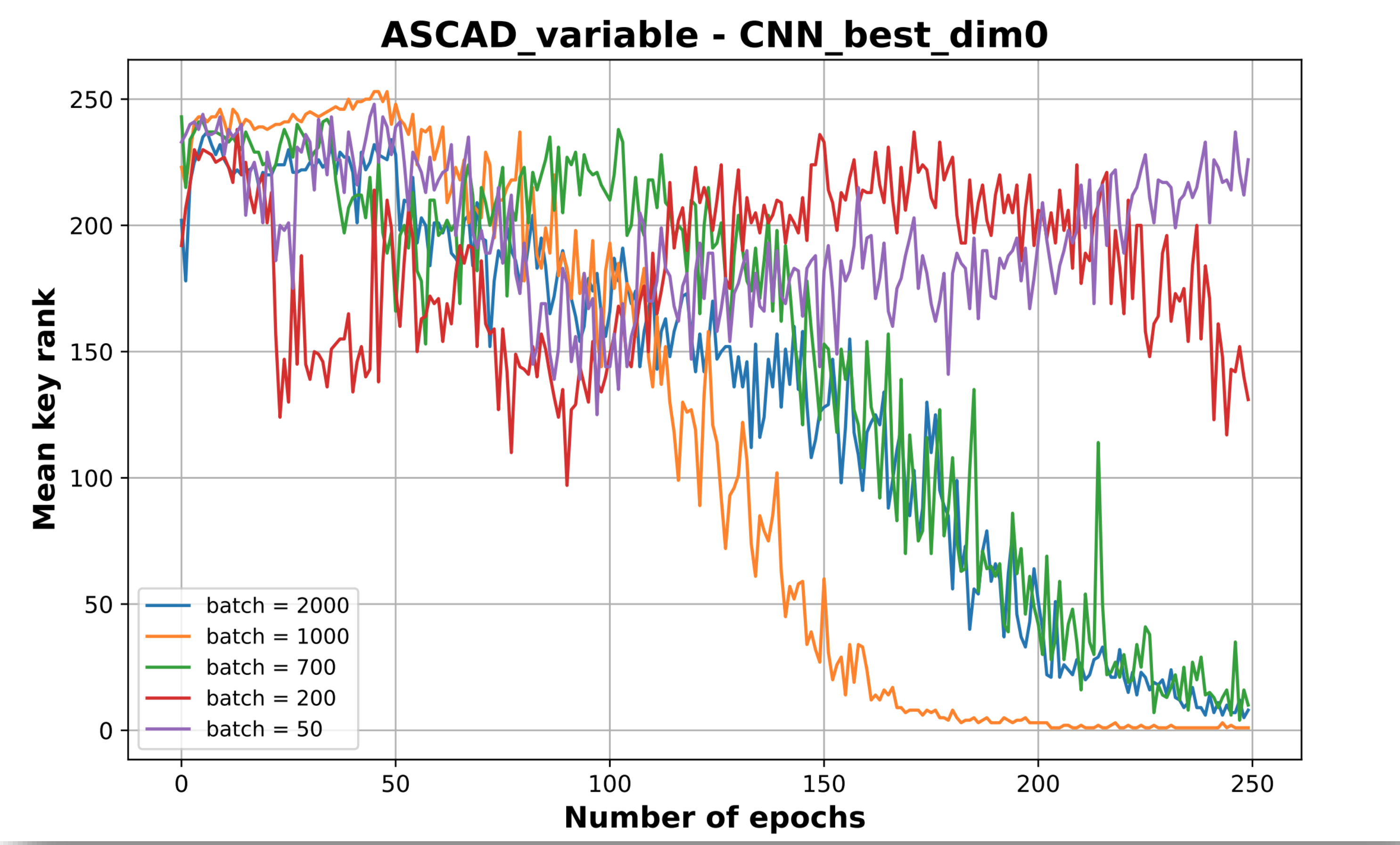


(b) ASCAD_desync100, CNN_zaid on dim0 and dim1 with SMOTE, HW labelling

[1] Wouters, L., Arribas, V., Gierlichs, B., Preneel, B.: Revisiting a methodology for efficient CNN architectures in profiling attacks. IACR Transactions on Cryptographic Hardware and Embedded Systems pp. 147–168 (2020)

ASCAD – WITH VARYING KEYS DURING PROFILING

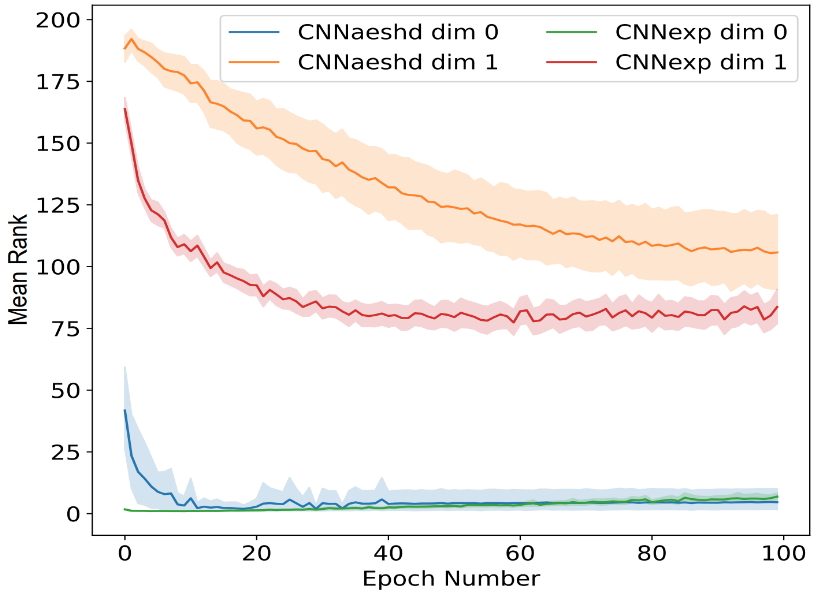
Technique	Approx. number of traces	Comment
Dim 1 with ID [1]	1000	Works 37 out of 60 times
Dim 1 with HW	N/A	
Dim 1 with HW + SMOTE	N/A	
Dim 0 with HW	10’000	



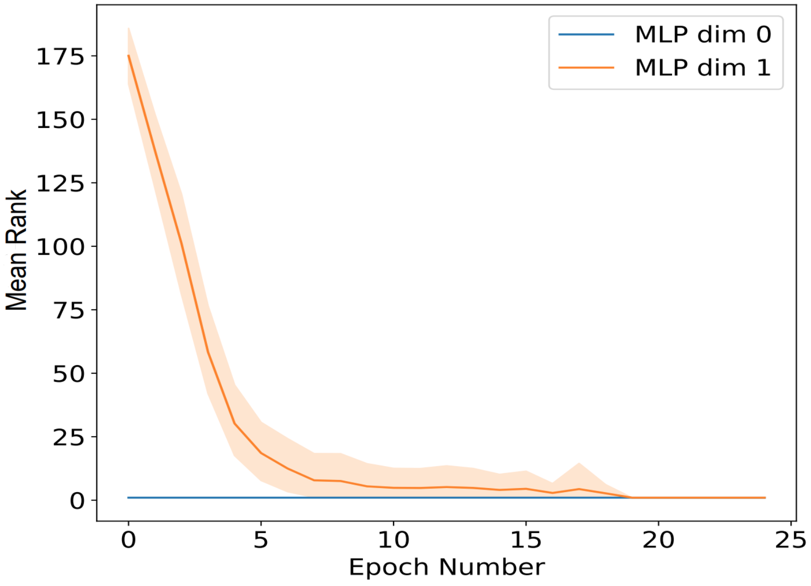
[1] Egger, M., Schamberger, T., Tebelmann, L., Lippert, F., Sigl, G.: A second look at the ascad databases. In: International Workshop on Constructive Side-Channel Analysis and Secure Design. pp. 75–99. Springer (2022)

OTHER DATASETS – AES_HD & DPACONTEST

- Compares against own implementation of dimension 1 VS 0, not State-of-the-Art



(a) AES_HD dataset.



(b) DPAContestv4.2 dataset.

Attacking various HW-labelled datasets 50 times, comparing dim0 and dim1

BONUS: UNPROFILED ATTACKS – AES_HD

Unprofiled performance of the CNNexp model on the AES_HD dataset
using a HD labelling with 15 epochs and over 10 attacks for each dimension.

Attack Number	1	2	3	4	5	6	7	8	9	10
dim0 Key Rank	1	1	2	10	6	2	1	1	14	1
dim1 Key Rank	126	2	80	3	18	168	1	66	5	58

Technique	Key Rank 1 Success Rate	Key Rank 20 Success Rate
Dim 1 with HD	10%	50%
Dim 0 with HD	50%	100%

csem

CONCLUSION – DIMENSION 0

- Straightforward implementation
- Not generalizable to other applications
- Separate class training
  - With an adaptive optimizer, and exponential decay
  - No Inter-class bias
- “Comparing” input traces with each other
- Varying global ranking trace-weights

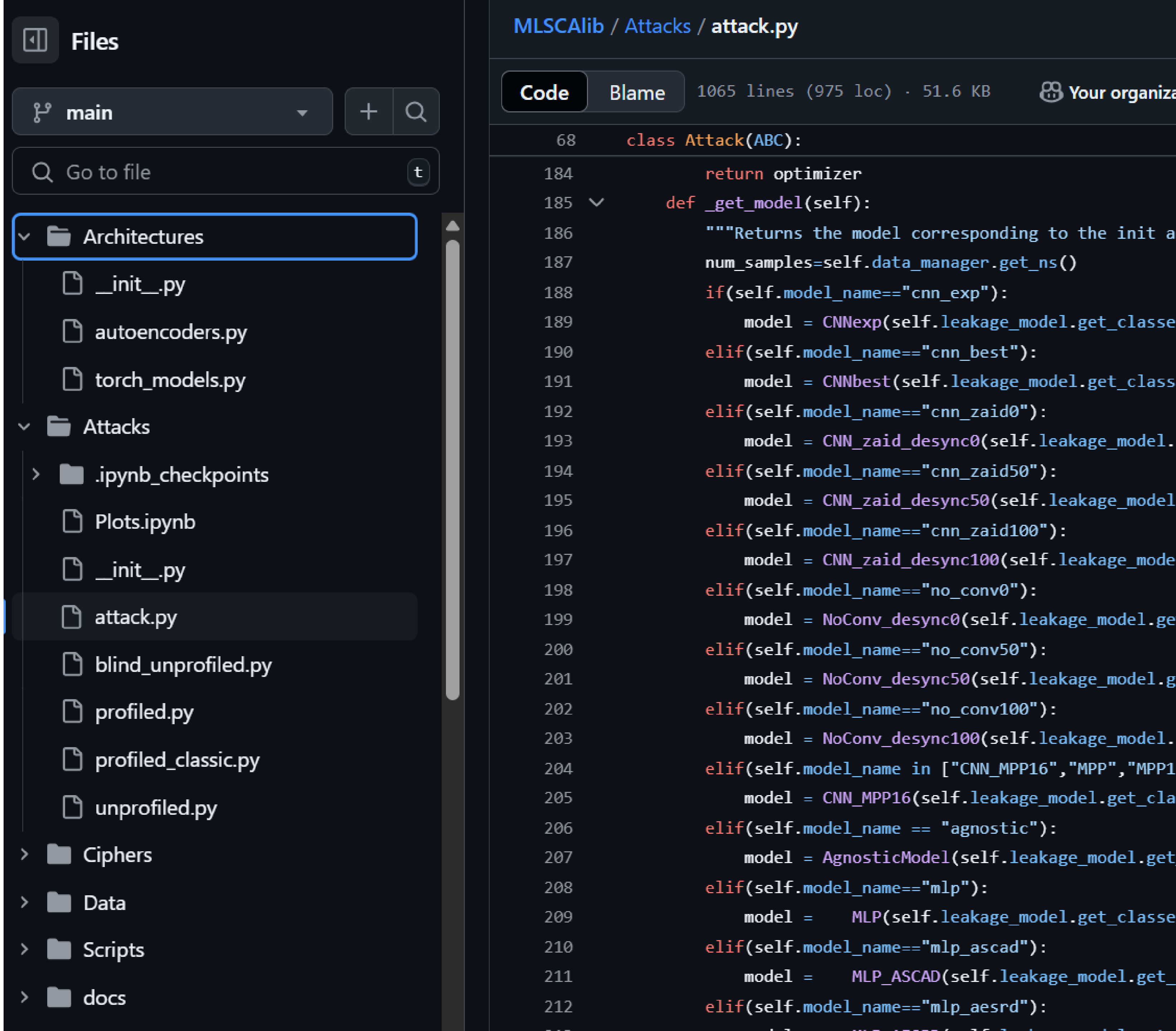
THE MLSCALIB:
A LIB FOR SCA & ML

INTRODUCING THE MLSCALIB

- Usable via command line & as package
- Implements dozens of ML publications for SCA
- Detailed documentation
- https://github.com/csem/MLSCAlib

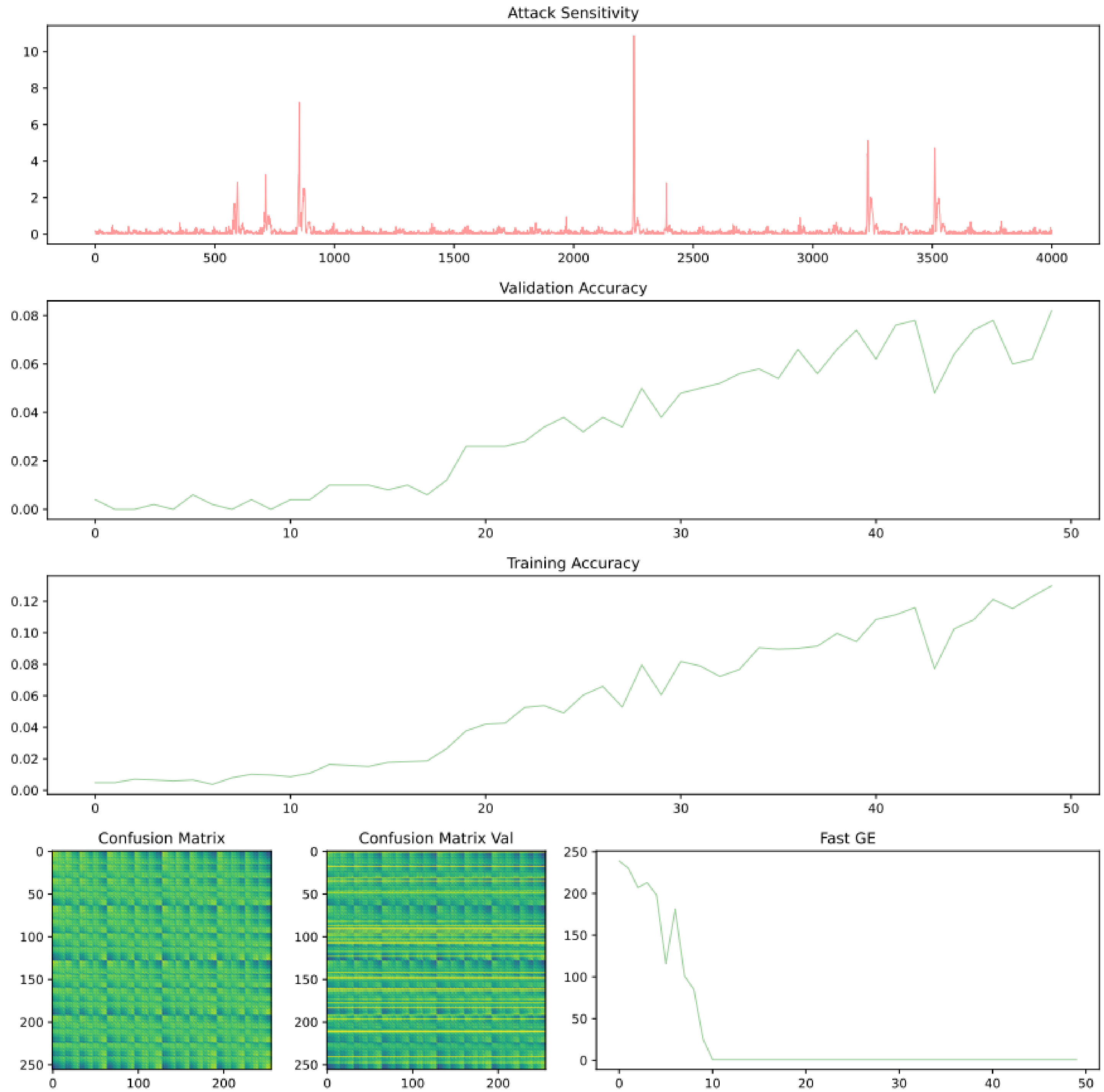
MLSCALIB MODELS

- 30 PyTorch models
- 1 autoencoder



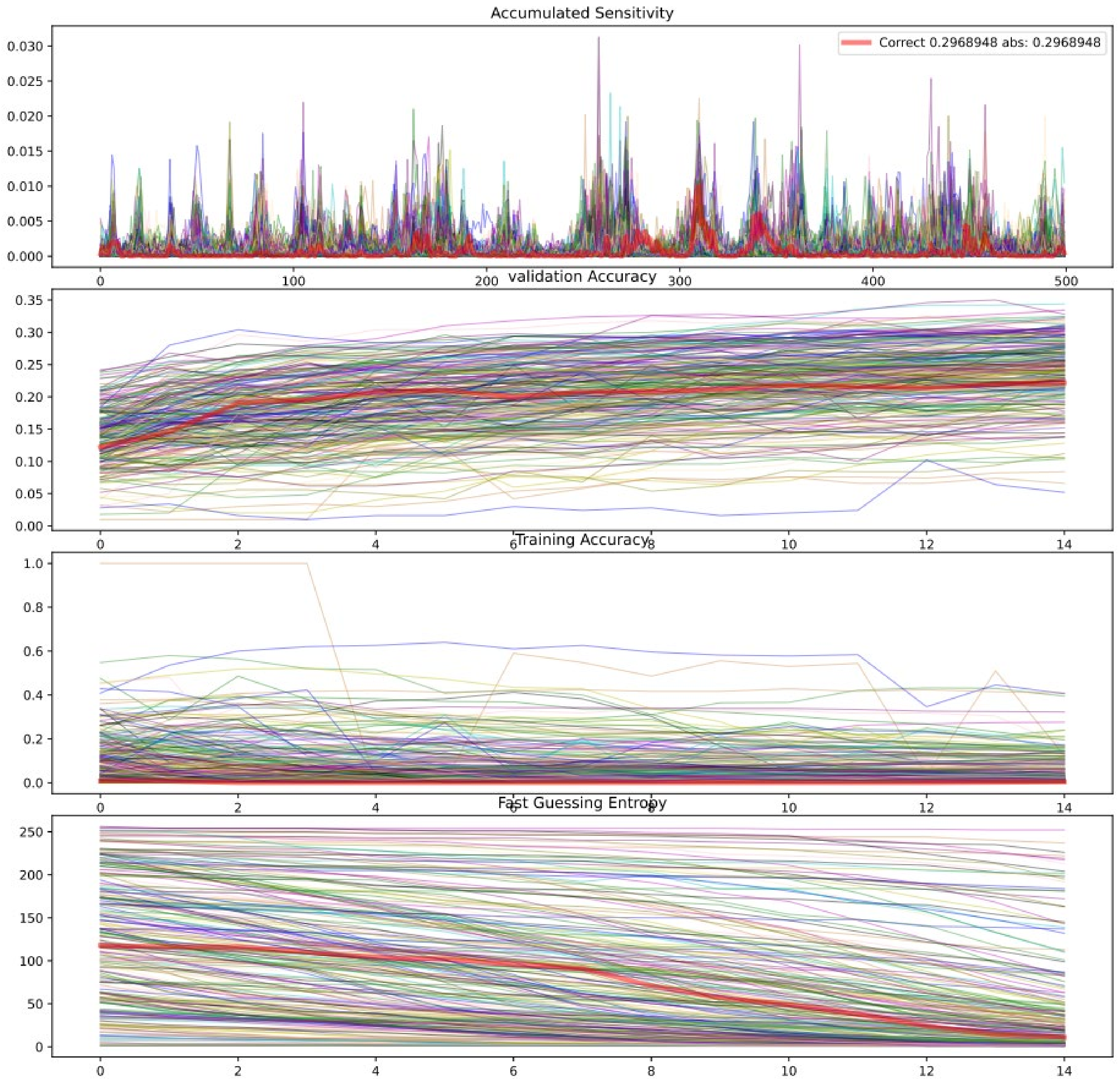
MLSCALIB PLOTS

- Gradient Visualization
- Accuracies
- Confusion
- Fast Guessing Entropy



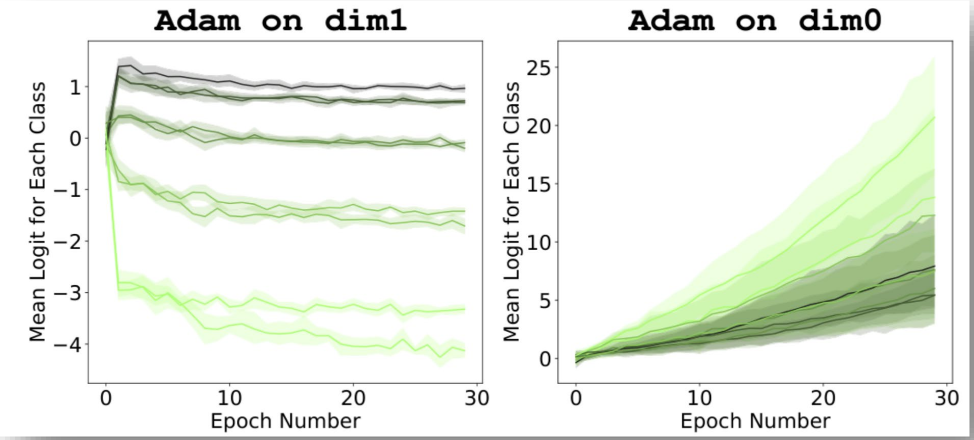
MLSCALIB PLOTS

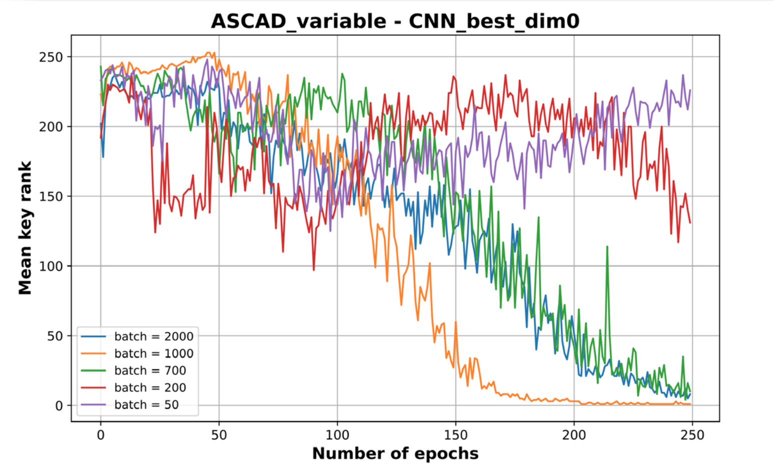
- Unprofiled attacks, for each guess:
  - Gradient visualization
  - Accuracies
  - Fast Guessing Entropy

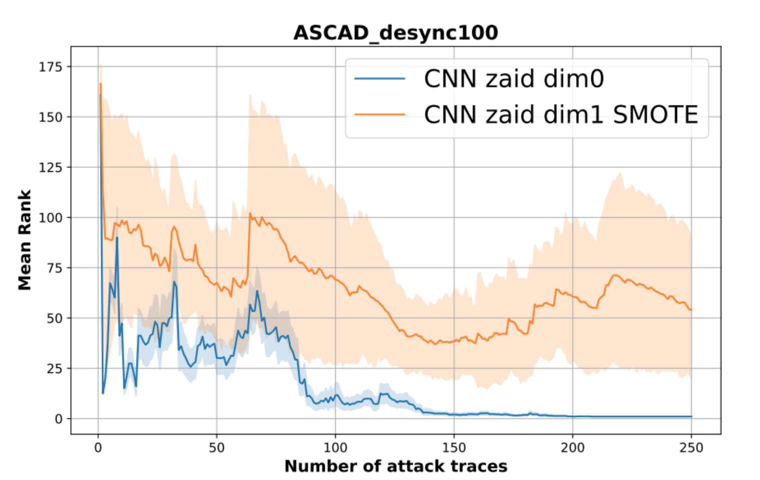


MLSCALIB BENCHMARKS

- Compares attacks
- 2 types of x-axis







csem

csem

FACING THE CHALLENGES OF OUR TIME