2023 Feb 2

Romanas Munovas, s4004981

The Importance of Filters:

Using Shapley Value Pruning to Optimize Convolutional NNs

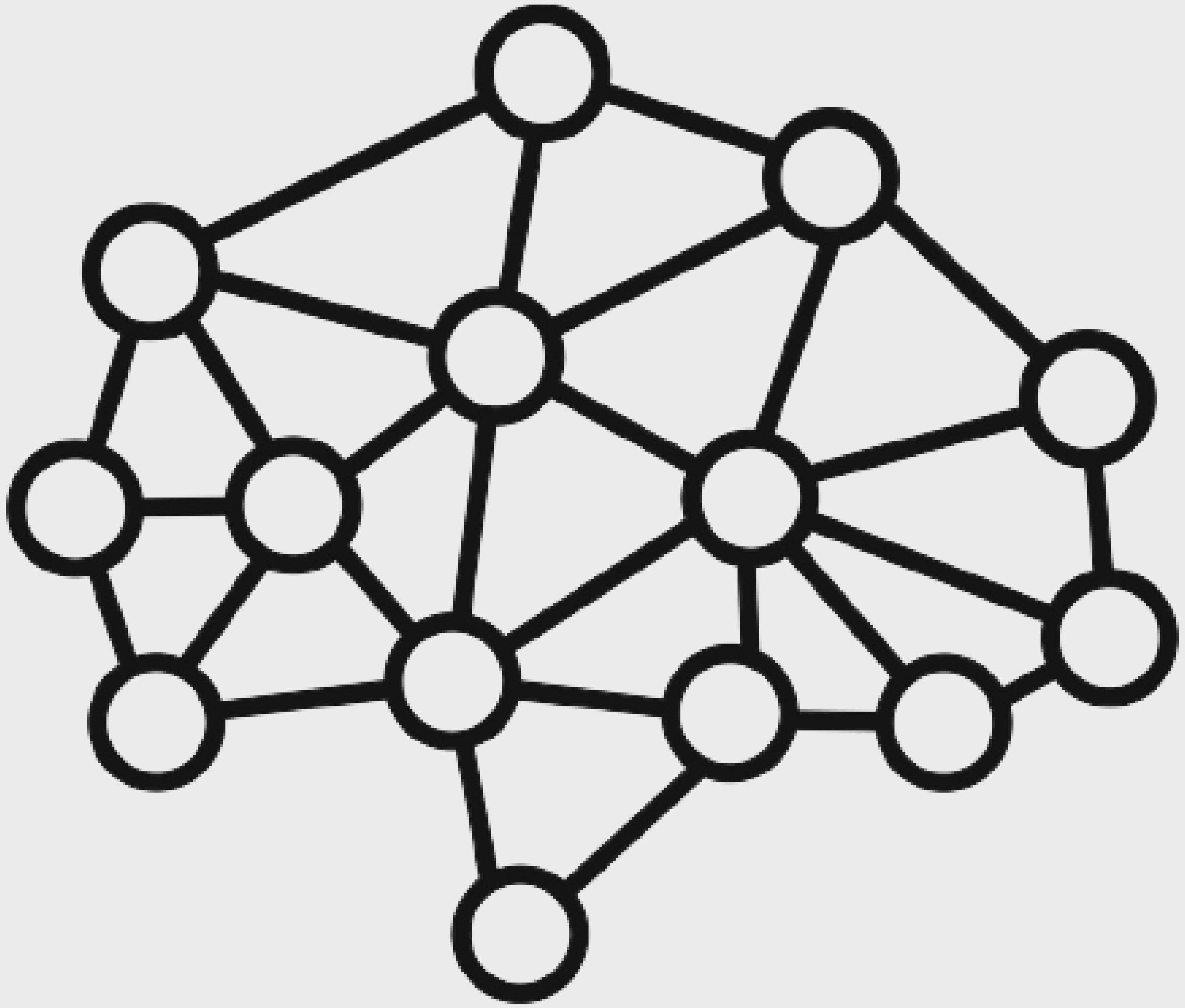
Bachelor's Project Presentation

/ prun / - prune

Prune

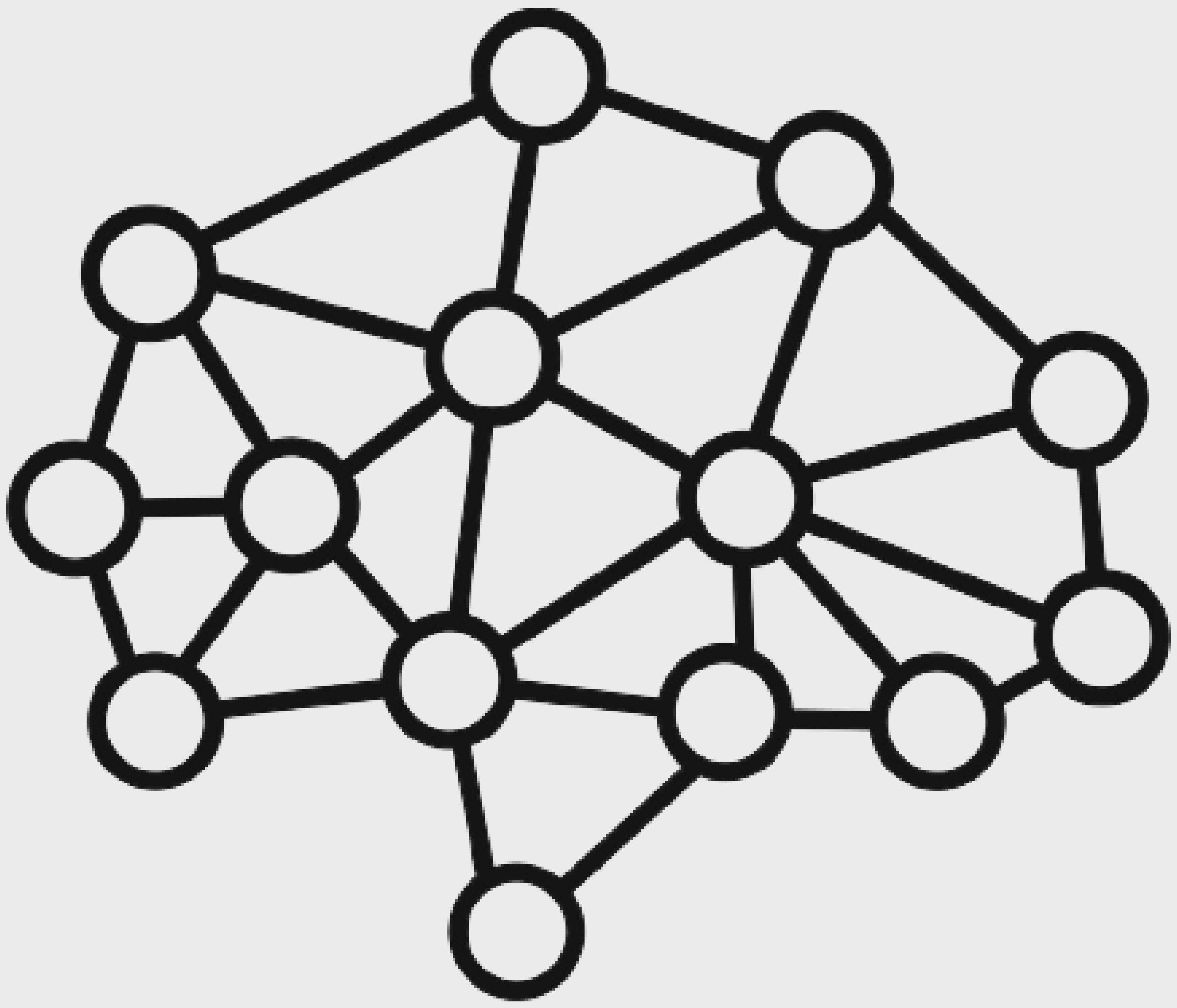
(verb) weed out unwanted or unnecessary things

What about neural networks?



Neural Network Pruning

- Unstructured (weights)
- Structured (filters/layers)



Structured Pruning

Removing **whole** filteres/layers

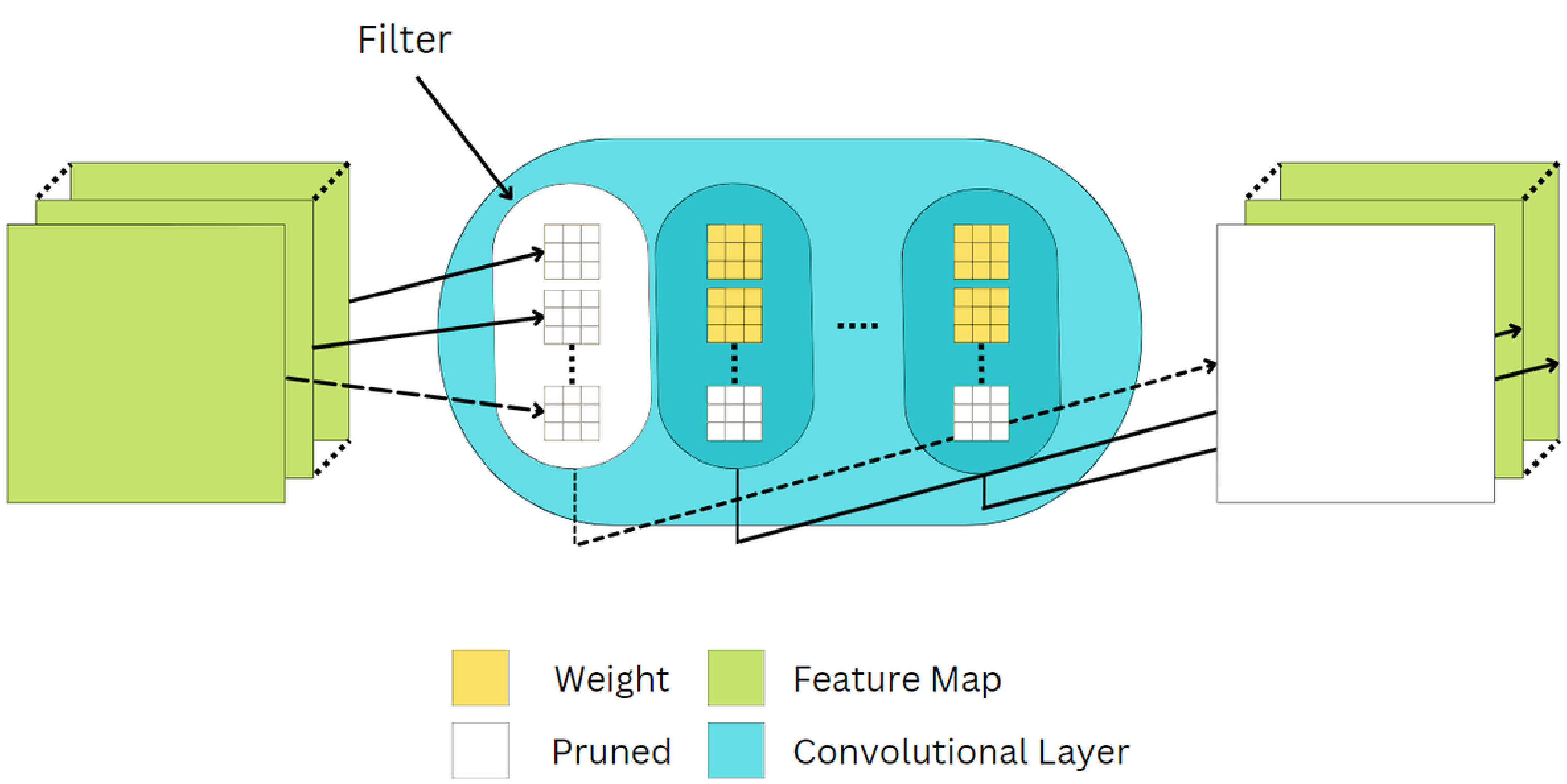
PRUNING FILTERS FOR EFFICIENT CONVNETS

Hao Li* University of Maryland haoli@cs.umd.edu

Asim Kadav NEC Labs America asim@nec-labs.com

Igor Durdanovic NEC Labs America igord@nec-labs.com

- Eliminates the feature maps



- Smaller in size
- Decrease computational costs
- Reduce inference time

Pruning Procedure

Train a neural network

Prune based on importance score

Fine-tune

Pruning Criteria

- Saliency of parameters
- Average Percentage of Zeros
- Batch normalization scaling factors
- Magnitude for each channel of filters
- Geometric mean

Studying the Plasticity in Deep Convolutional Neural Networks using Random Pruning

Deepak Mittal · Shweta Bhardwaj · Mitesh M. Khapra · Balaraman Ravindran

2018

- Pruned 25-50% of the filters randomly from deep CNNs
- Yielded same performance as state-of-art pruning methods

Importance score - redundant metric?

Shapley

Shapley
Literature Findings

Shapley Values as a Principled Metric for Structured Network Pruning

2020

- Introduces Shapley values for pruning
- Shows that Shapley criteria superior to other criteria
  - Game-theoretic foundation
  - Outperformed other pruning criteria
  - Works in low-data regimes, when fine-tuning is either unavailable or ineffective

Shapley

Original Paper (1952)
Literature Findings

A VALUE FOR n-PERSON GAMES

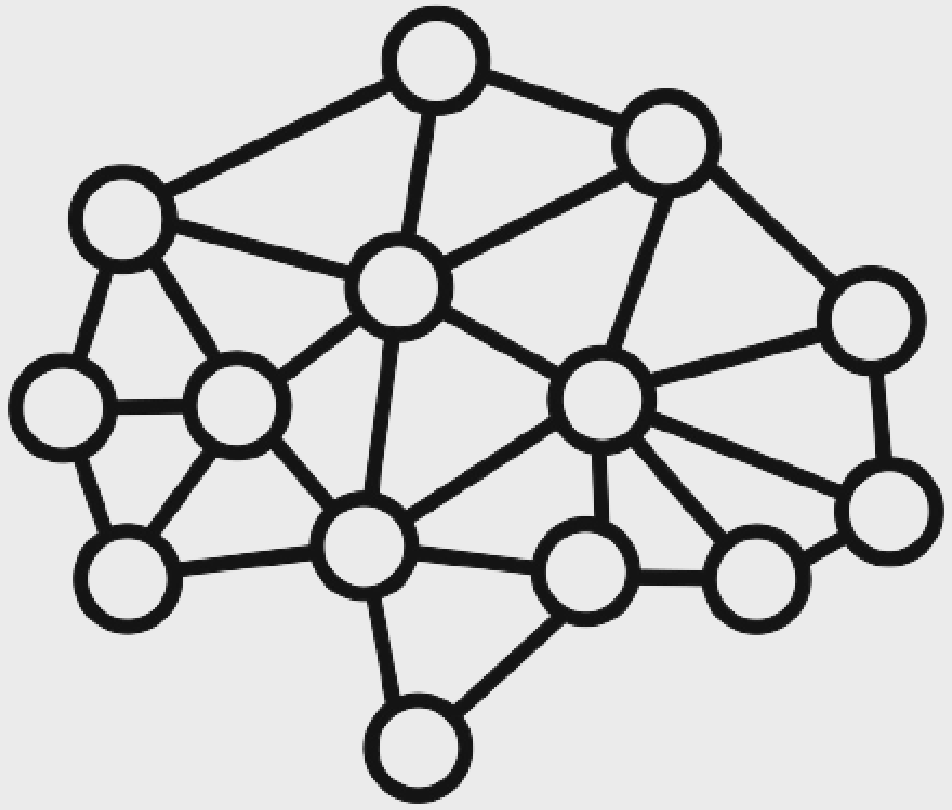
L. S. Shapley

- Proposed a way to compute the contribution of each player in a cooperative game.

Shapley

Neural Networks

- A CNN can be viewed as a cooperative game with the aim to minimize loss
    - The filters are the players, each contributing towards the minimization goal



Shapley

Marginal Contribution

- **Marginal Contribution: difference between importance of all features S with and without i**

$$m_x(S, i) := v_x(S) - v_x(S \setminus \{i\})$$

S Subset of all players S ⊆ P

i Player from subset S

v_x Function calculating the score when set of players S participate

Shapley

Shapley Value

Shapley value: average of marginal contribution distributions

$$\phi_x(i) := \frac{1}{n!} \sum_{S \subseteq P \setminus \{i\}} (|S|!(n - |S| - 1)!) m_x(S, i)$$

$$12! = 479001600$$

Pruning

Methodology

1. **Create a set of all possible filter permutations (coalitions)**
    a. **If not possible, use Monte Carlo sampling**

2. Calculate the loss of a single forward pass (all filters)

3. For each filter permutation, mask the given filters and calculate the marginal contributions of the masked filters.

4. Shapley value for filter i is the aggregated average of all marginal contributions.

Setup

1

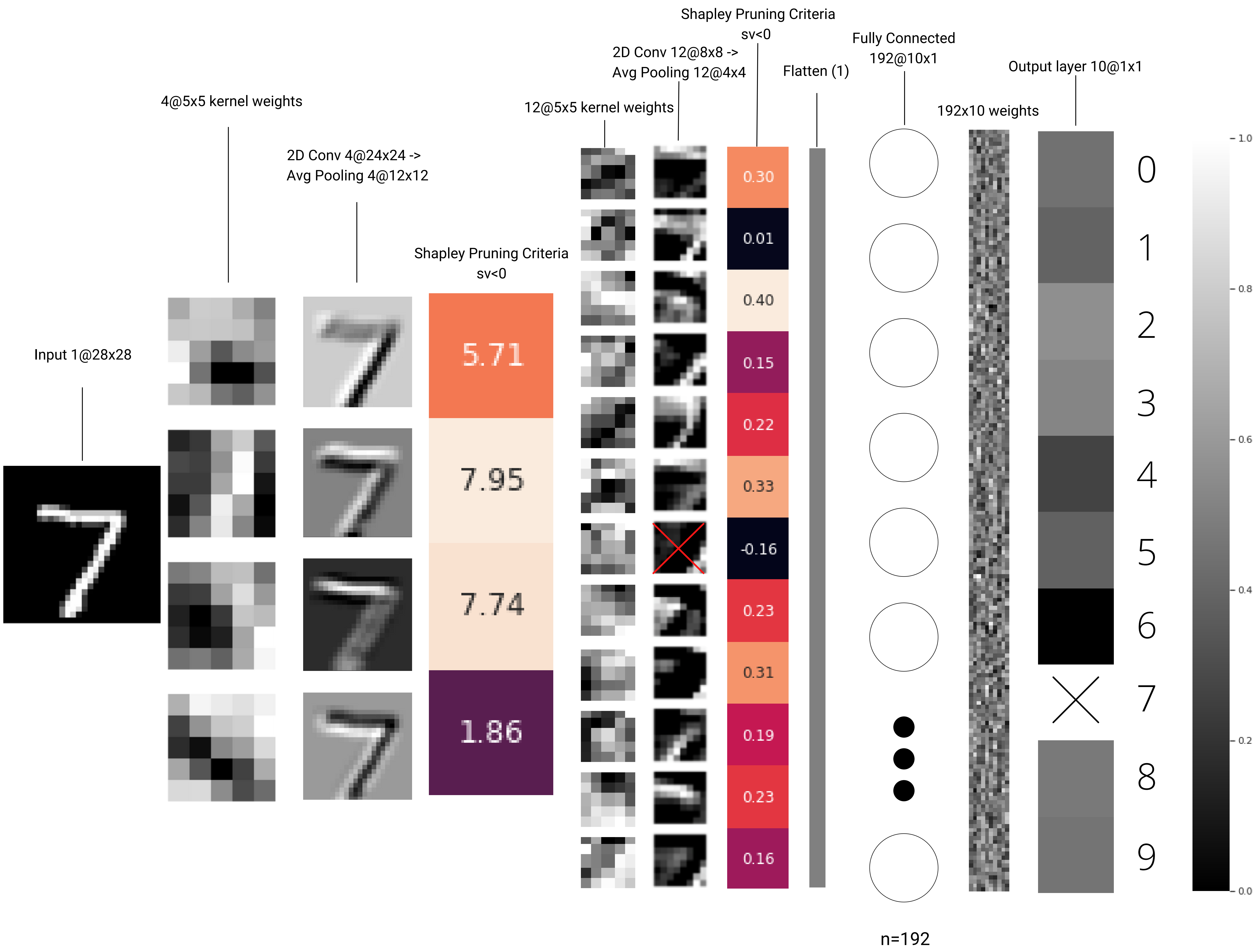
LeNet-1
Replicated

2

98.4% accuracy
(On MNIST)

3

Cross-entropy
Loss as the
criterion



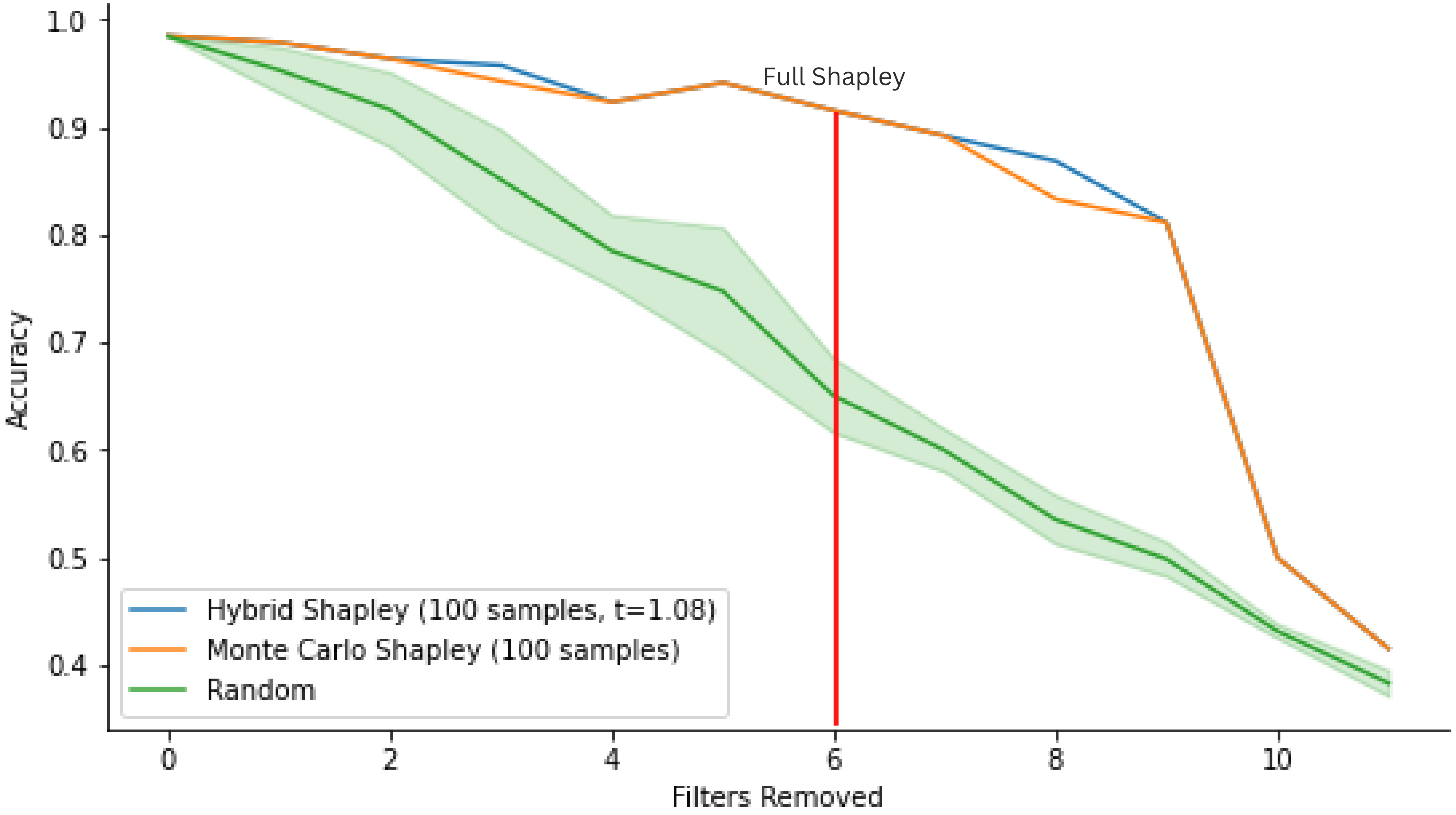
Shapley Pruning Criteria
sv<0
2D Conv 12@8x8 ->
Avg Pooling 12@4x4
Flatten (1)
Fully Connected
192@10x1
Output layer 10@1x1
4@5x5 kernel weights
12@5x5 kernel weights
192x10 weights
2D Conv 4@24x24 ->
Avg Pooling 4@12x12
Shapley Pruning Criteria
sv<0
Input 1@28x28
5.71
7.95
7.74
1.86
0.30
0.01
0.40
0.15
0.22
0.33
-0.16
0.23
0.31
0.19
0.23
0.16
0
1
2
3
4
5
6
7
8
9
1.0
0.8
0.6
0.4
0.2
0.0
n=192

Pruning

Quantitative Results

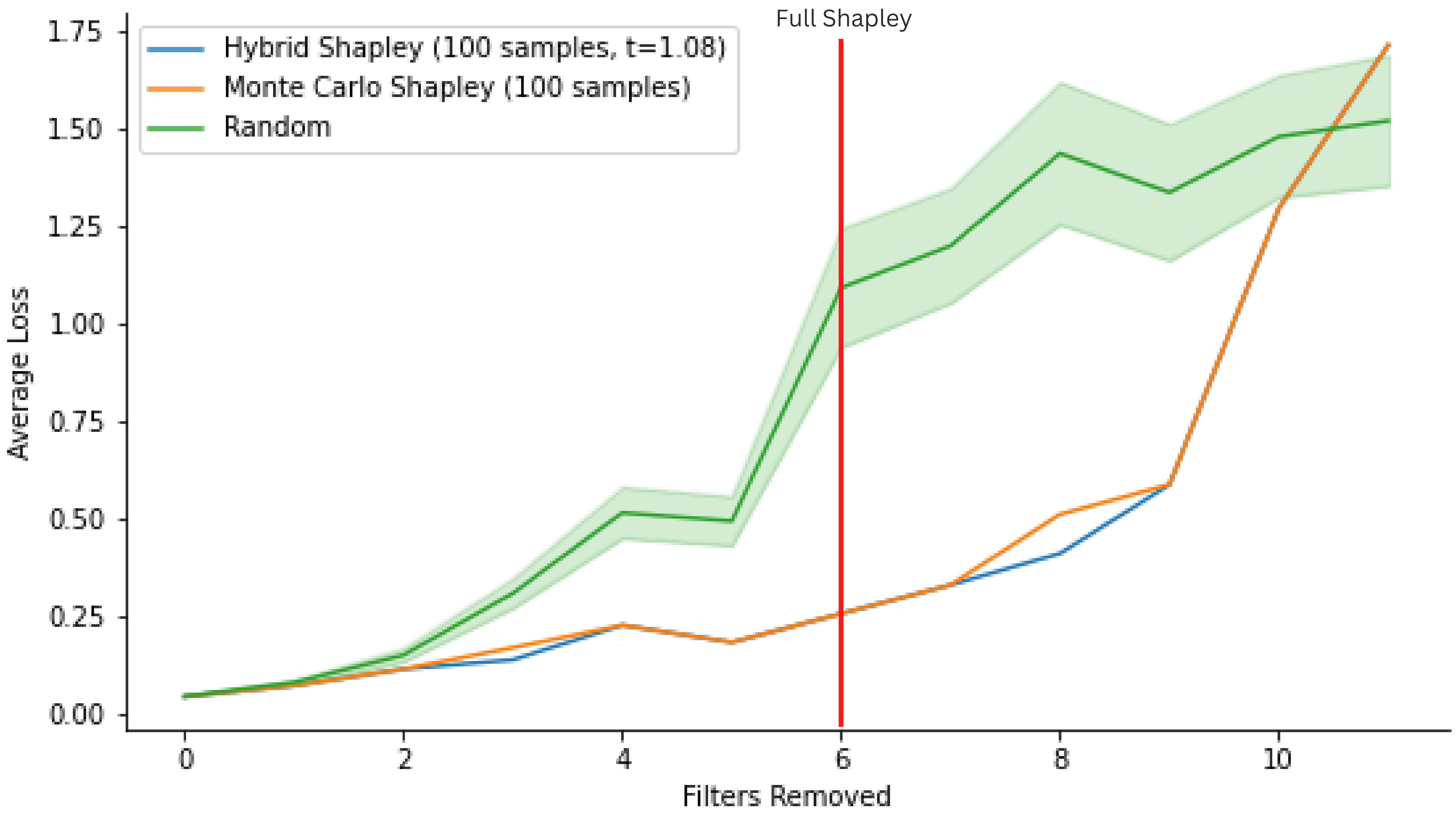
Accuracy (Test Data, no fine-tuning)

- Averaged over 3 trials



Loss (Test Data, no fine-tuning)

- Averaged over 3 trials



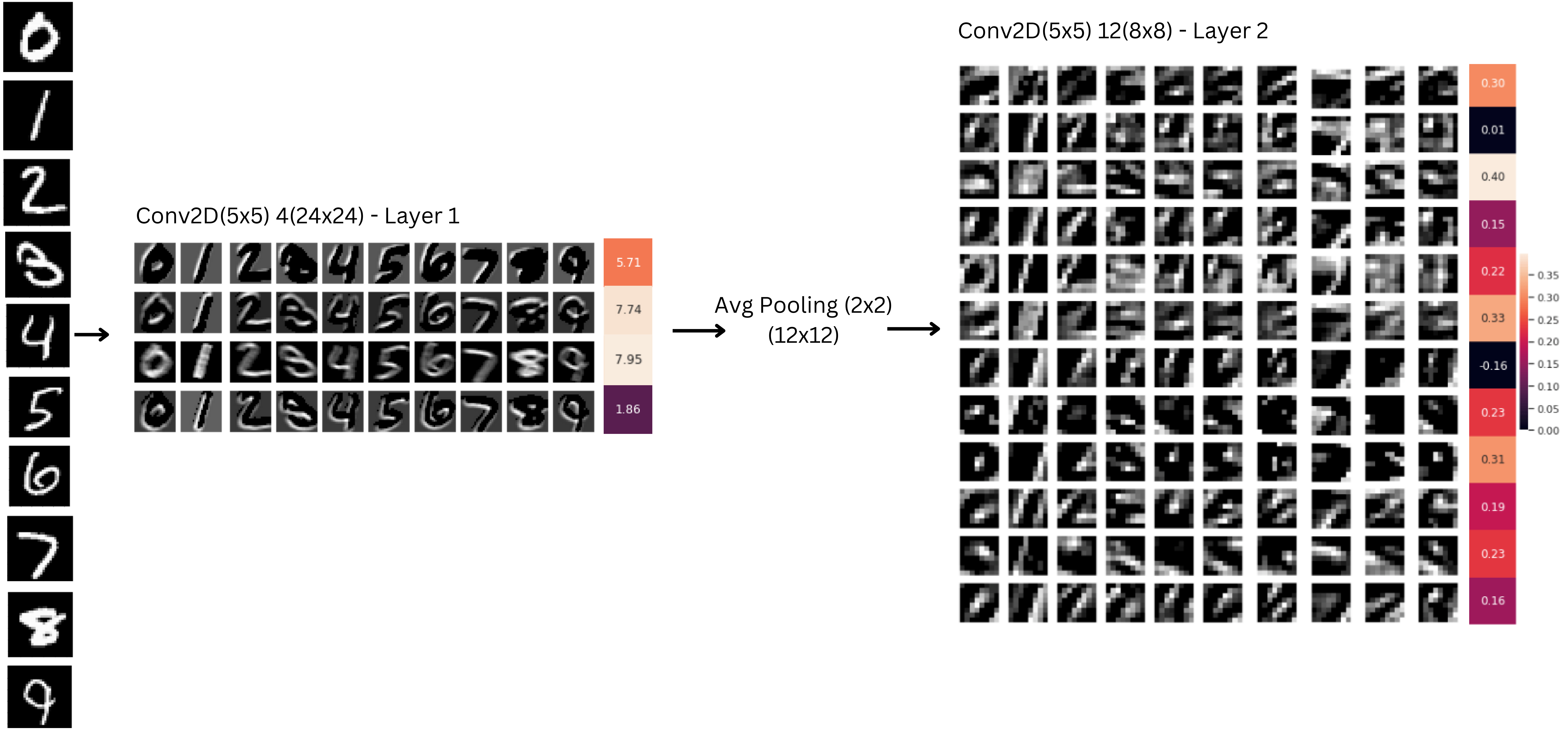
Pruning

Quantitative Results (AUC) Loss

Area Under the Curve (AUC) Loss

	With fine-tuning	Without fine-tuning
Random	0.11±0.01	**0.83±0.05**
Shapley	0.08±0.00	**0.41±0.00**

What does the machine deem important?

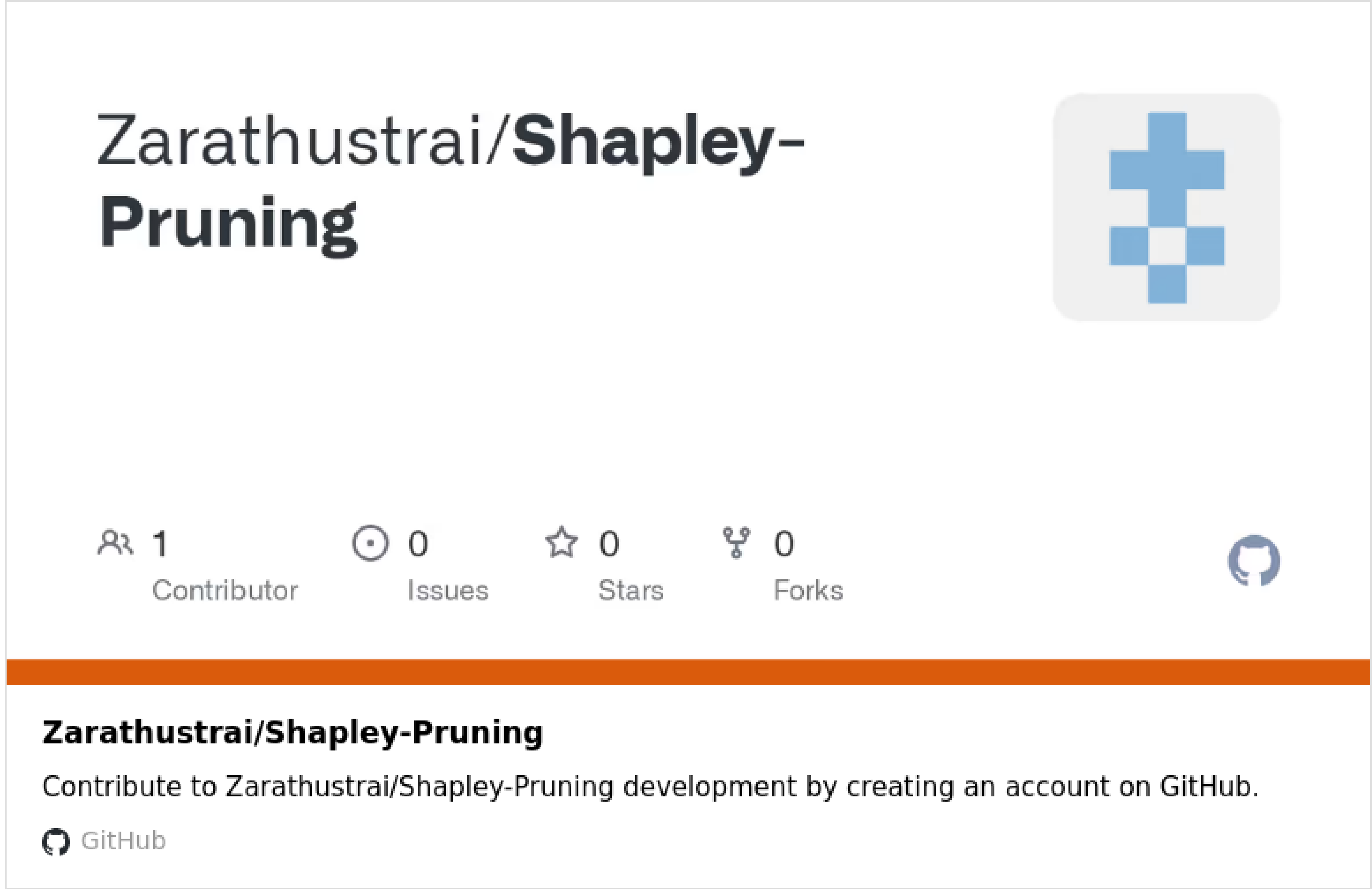

Conv2D(5x5) 4(24x24) - Layer 1
5.71
7.74
7.95
1.86
Avg Pooling (2x2)
(12x12)
Conv2D(5x5) 12(8x8) - Layer 2
0.30
0.01
0.40
0.15
0.22
0.33
-0.16
0.23
0.31
0.19
0.23
0.16
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

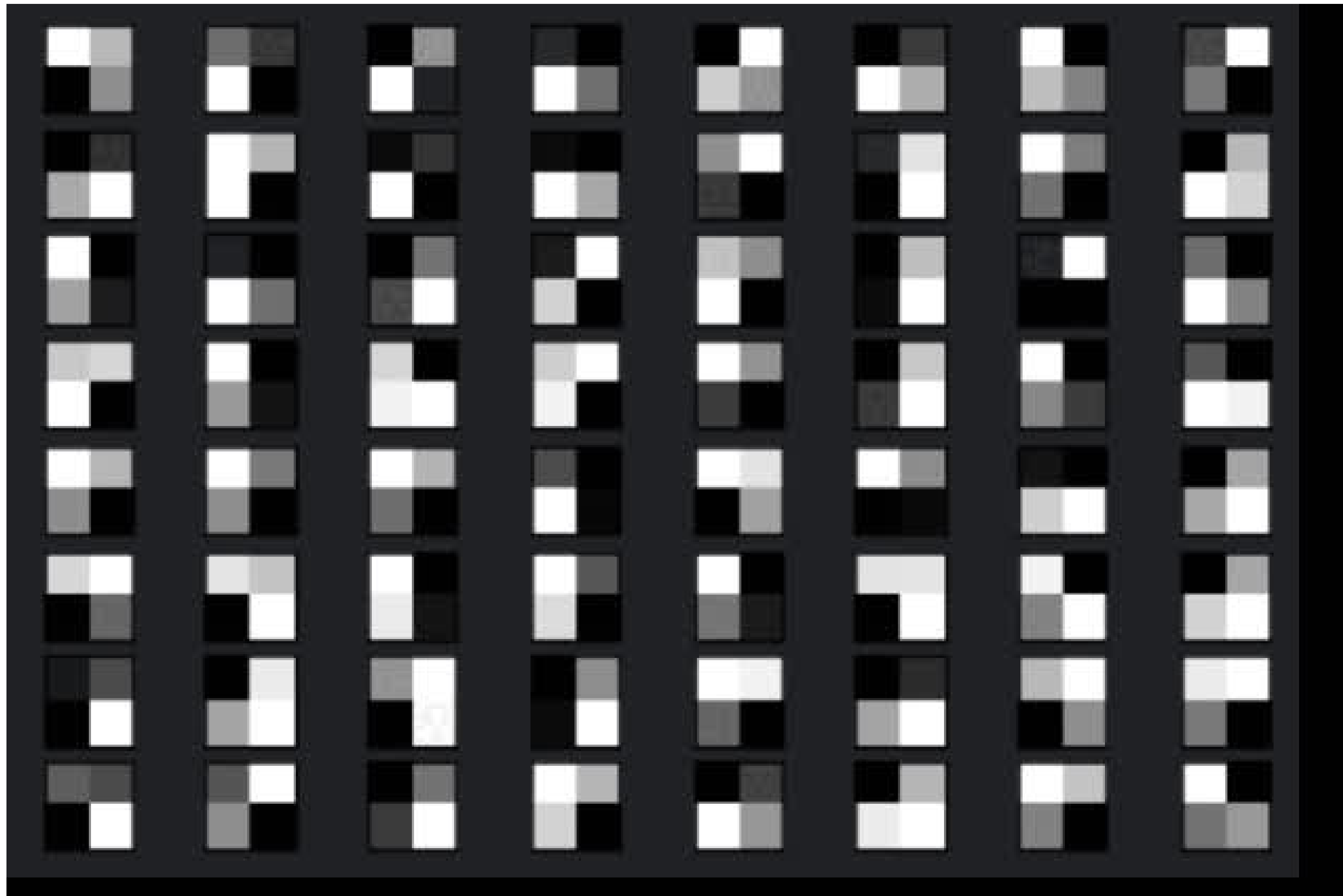
Qualitative Results

Feature Maps (Layer 1)

- Incomplete digits
- Leads to similarities between classes

Questions?

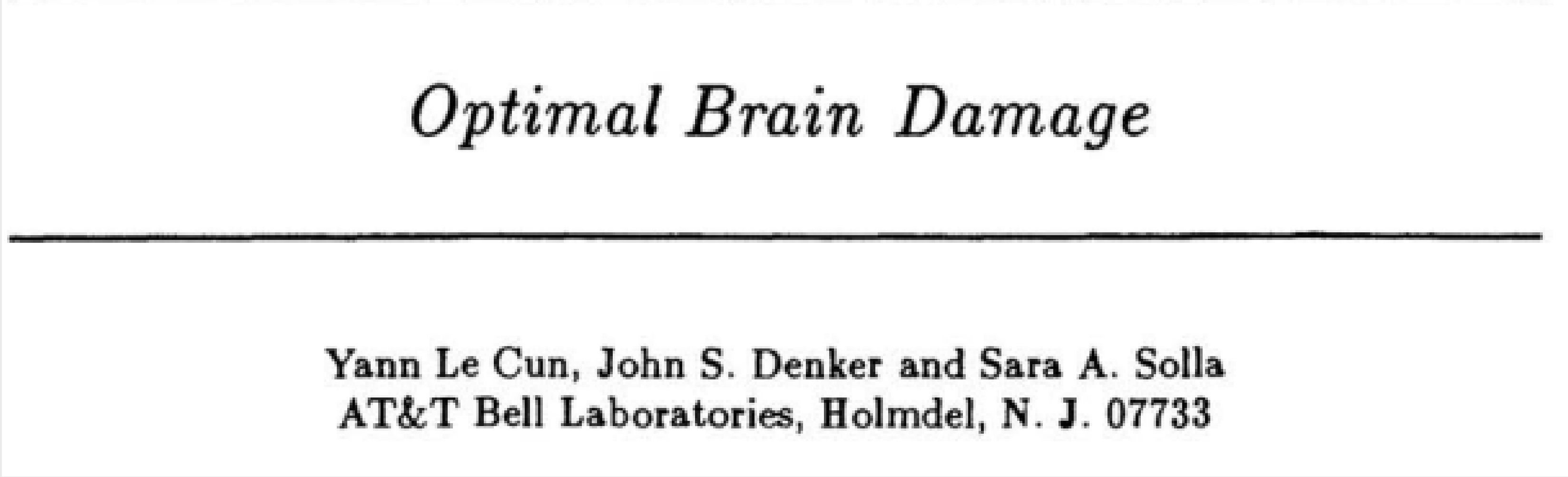

Zarathustrai/Shapley-Pruning
1 Contributor
0 Issues
0 Stars
0 Forks
Zarathustrai/Shapley-Pruning
Contribute to Zarathustrai/Shapley-Pruning development by creating an account on GitHub.
GitHub



Extra

1989

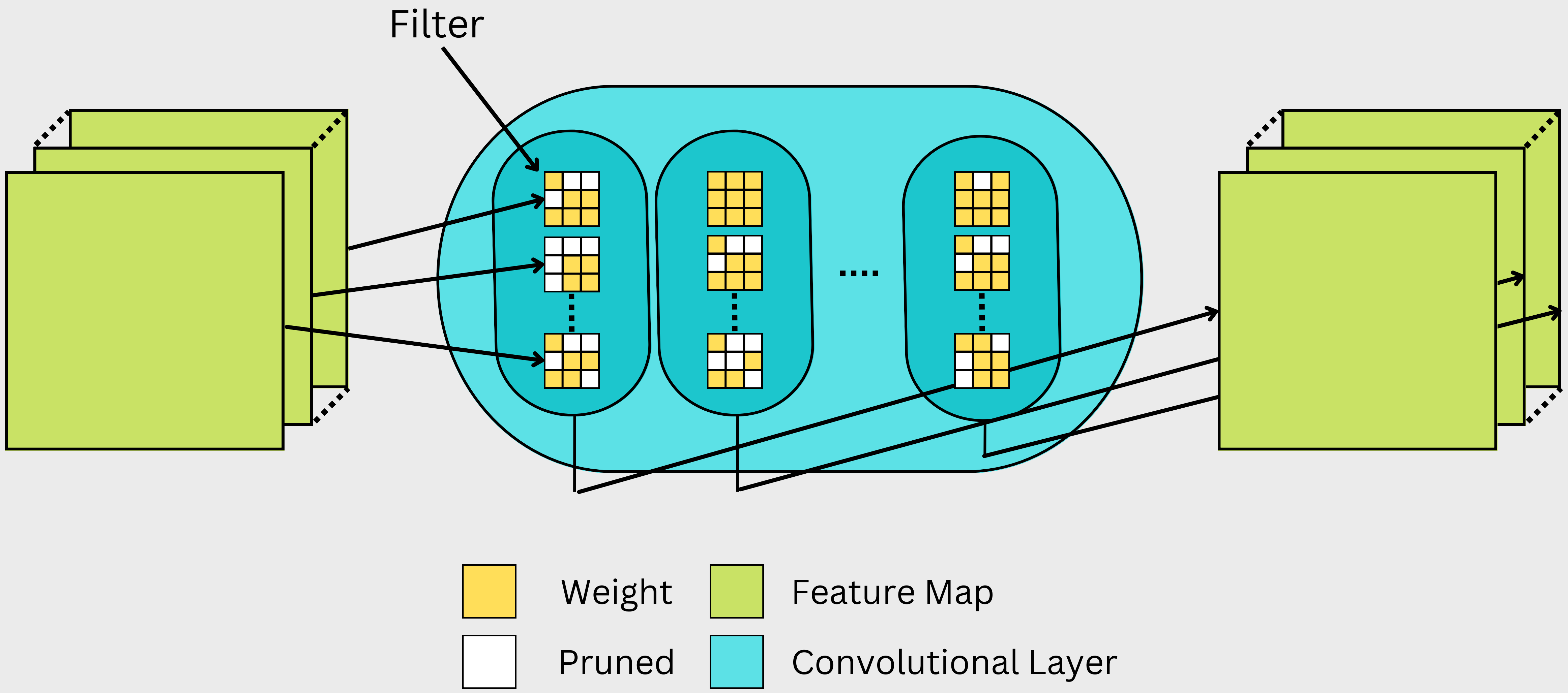
[Unstructured] Pruning introduced for the first time



Optimal Brain Damage

Yann Le Cun, John S. Denker and Sara A. Solla
AT&T Bell Laboratories, Holmdel, N. J. 07733

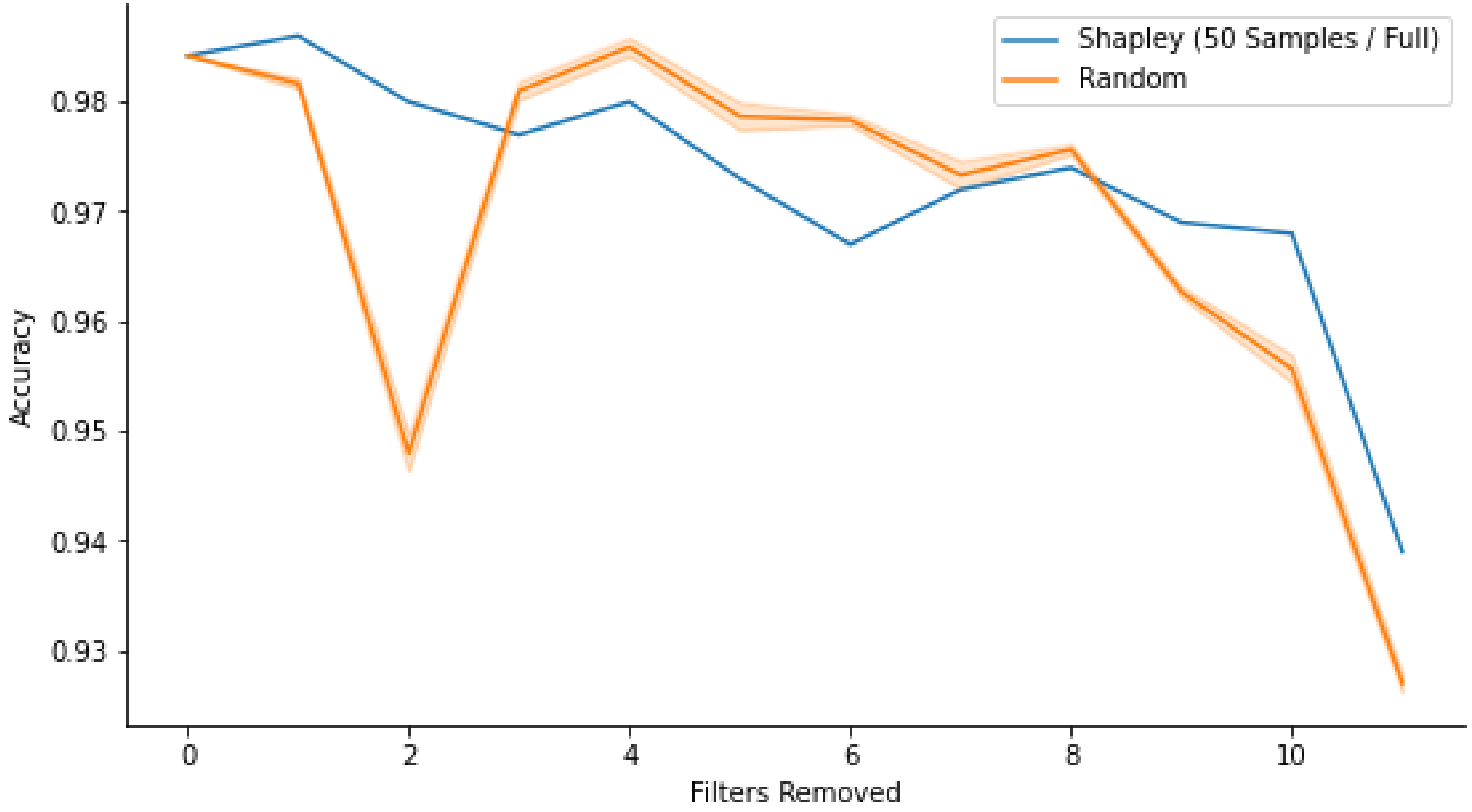
- Better generalization
- Fewer training examples required
- Improved speed of learning



Pruning

Quantitative Results
(normal-data regime)

Accuracy



Pruning

Quantitative Results (Comparison)

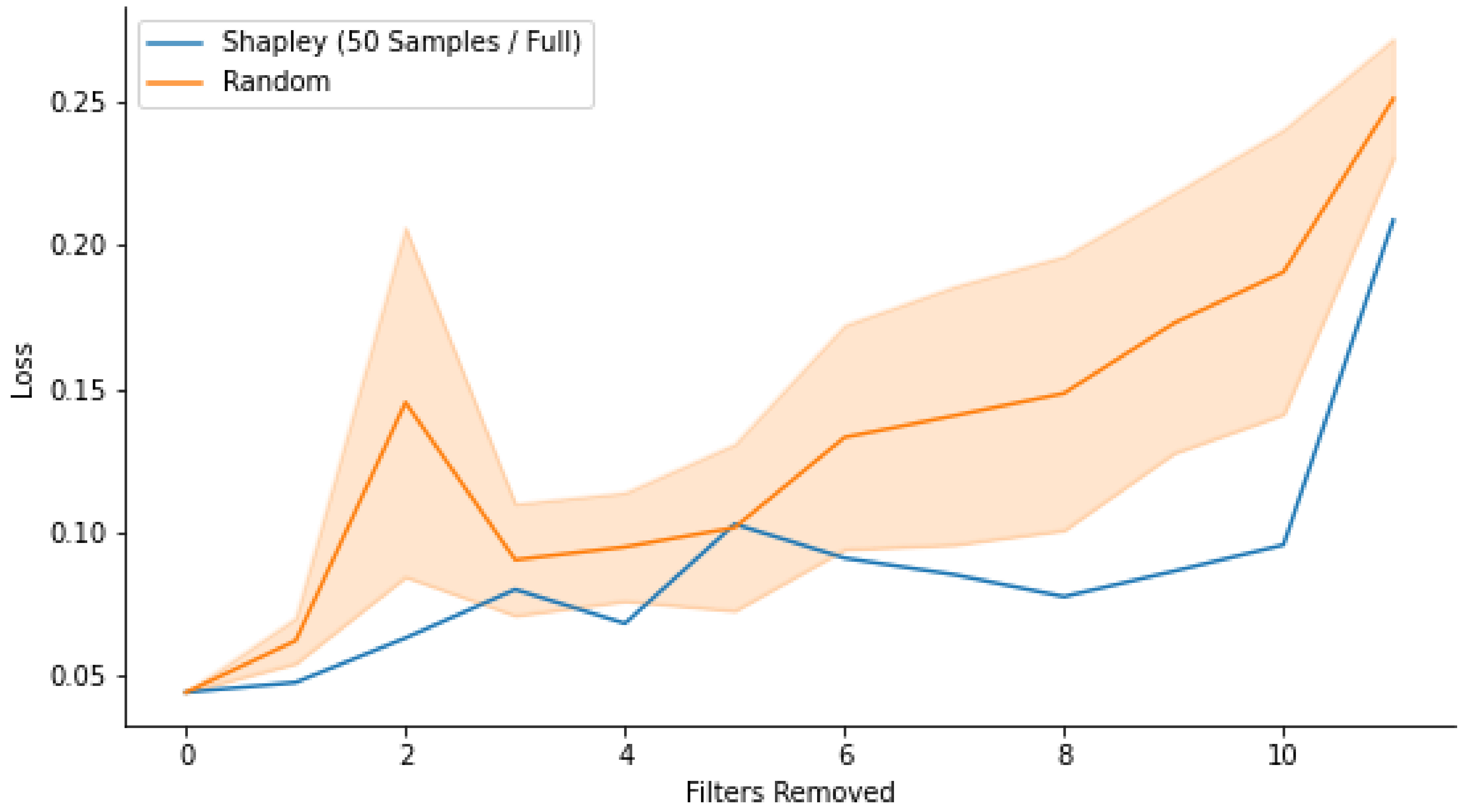
An Improvement!

	Accuracy (%)	Parameters
Original	98.3 (%)	3246
Pruned (1)	98.6 (%)	2985 (-8%)

Pruning

Quantitative Results
(normal-data regime)

Loss

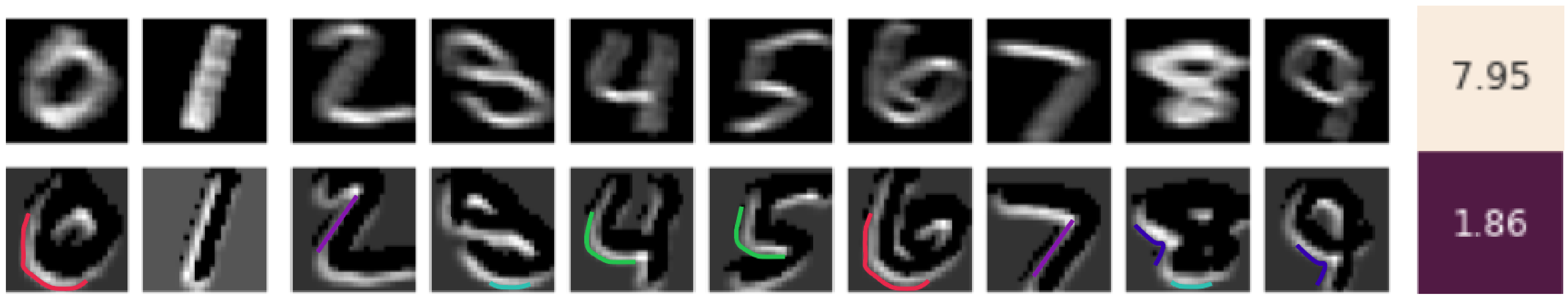


Pruning

Quantitative Results (Comparison)

An Improvement!

	Accuracy (%)	Parameters
Original	98.3 (%)	3246
Pruned (10)	96.8 (%)	636 (-81%)


7.95
1.86