Hyper-flexible Convolutional Neural Networks

Beyond Brittle: Forging Robust AI by Rethinking its Mathematical Core

Vagan Terziyan, Diana Malyk, Mariia Golovianko, Vladyslav Branytskyi

Faculty of Information Technology, University of Jyväskylä, Finland
Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Ukraine

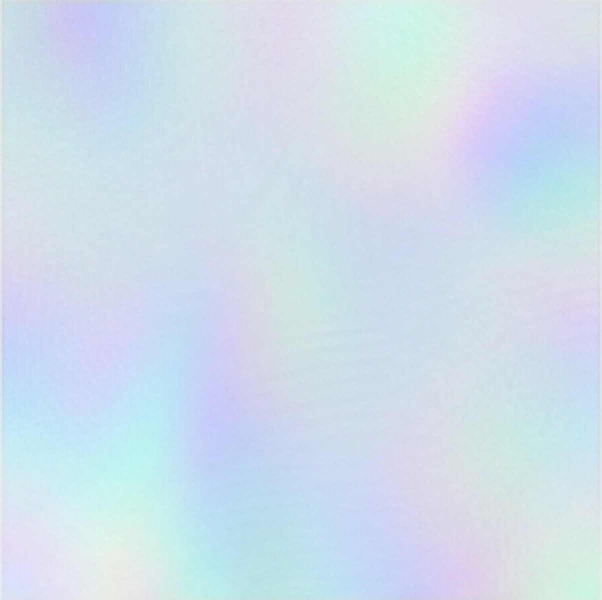
As published in *Neural Networks*, Volume 155, 2022.

Powerful, Yet Brittle: The Hidden Vulnerability of Standard CNNs

INPUT IMAGE

✅ Standard CNN classifies correctly with 98% confidence.

+



+ ADVERSARIAL PERTURBATION

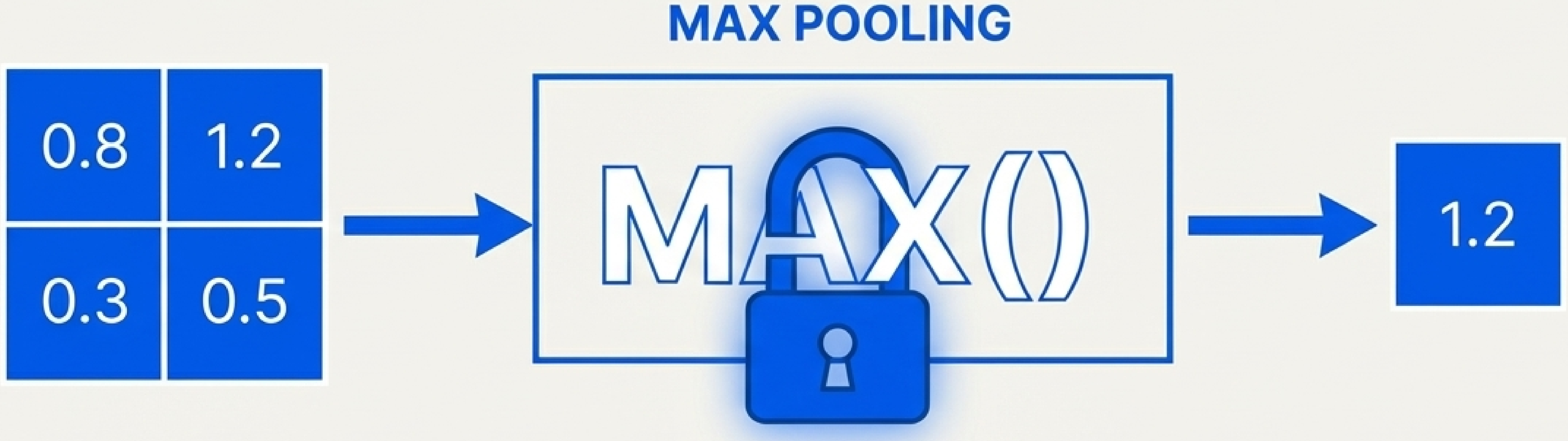
+

PERTURBED IMAGE

❌ Standard CNN misclassifies as ‘guacamole’ with 99% confidence.

State-of-the-art CNNs can be decisively fooled by tiny, human-imperceptible changes. **This fundamental lack of robustness is a critical barrier** to their thneir deployment in high-stakes, real-world scenarios.

The Root Cause is Mathematical Rigidity



0.8	1.2
0.3	0.5

The network learns *weights* for a *fixed* operation, but it cannot learn the *operation itself*.

Most CNN components—convolution, pooling, activation—rely on fixed mathematical functions. The network is highly tuned to the training data *for that specific function*, making it inflexible and prone to hidden overfitting. The mathematical essence of each component remains the same, regardless of the task.

The Solution: From Fixed Operations to Flexible, Learnable Functions

What if the network could learn the optimal mathematical operation for the task? We introduce two families of generalized mean functions that make this possible.

Generalized Lehmer Mean (GLM)

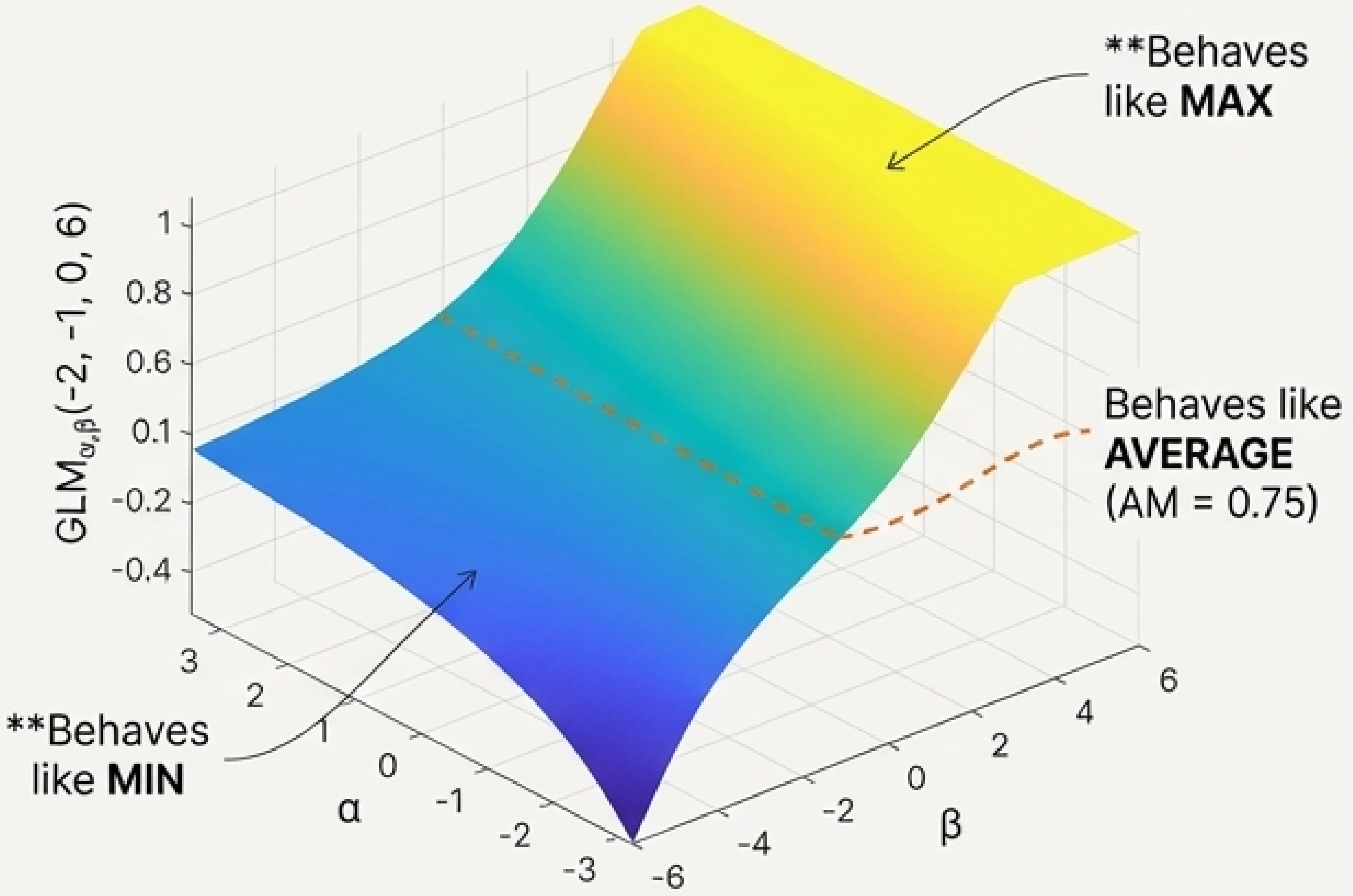
A two-parameter (α`, `β`) function that can smoothly transition between MIN, AVG, and MAX operations and a continuous spectrum of behaviors in between.

Generalized Power Mean (GPM)

A similar two-parameter (`$\gamma$`, `$\delta$`) function offering a rich and distinct spectrum of aggregation behaviors.

By making these parameters **trainable**, the network itself discovers the ideal mathematical behavior for each component, moving beyond fixed logic to achieve a new level of **adaptability.**

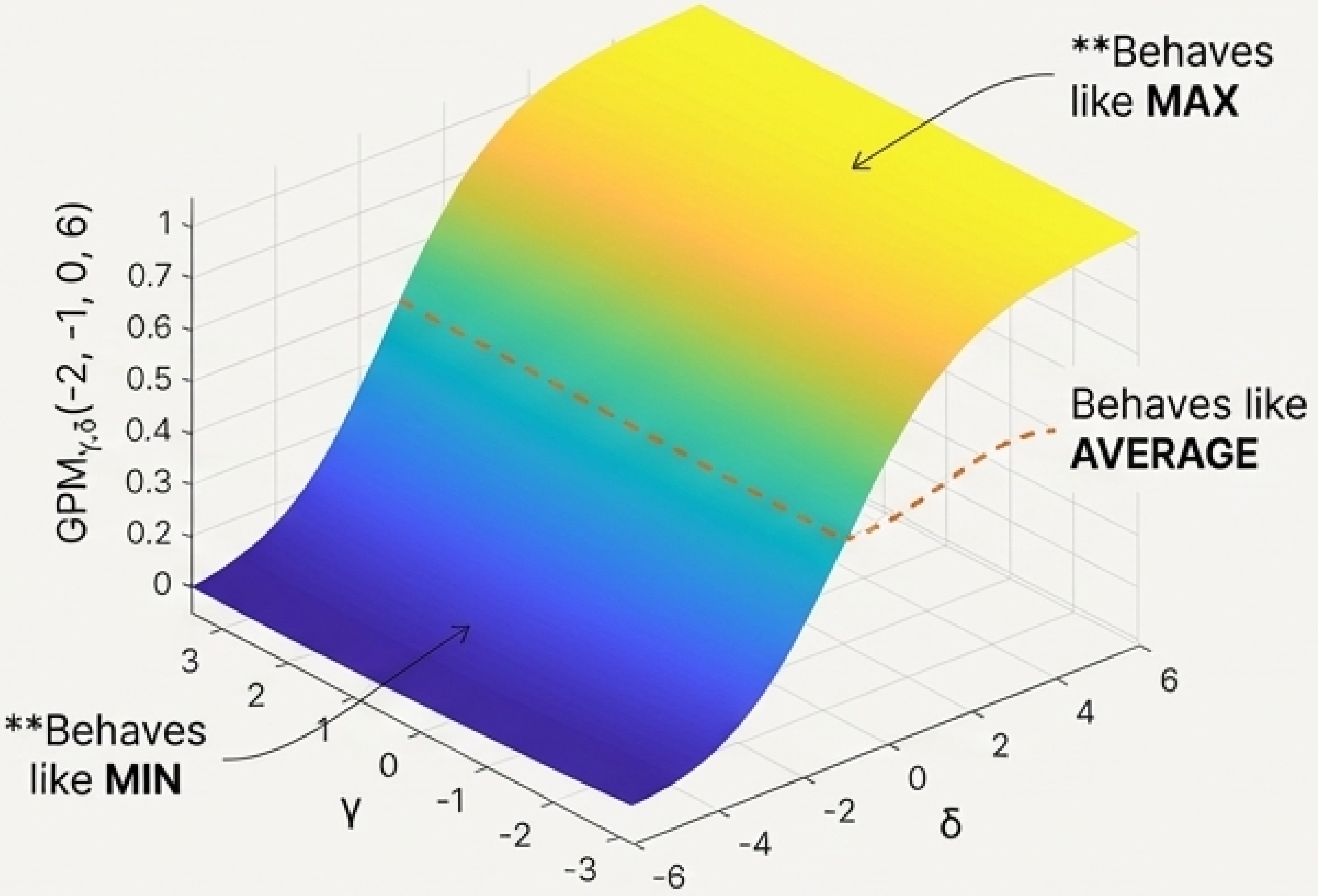
A Universe of Functions in a Single Equation



Behaves like **MAX

Behaves like **AVERAGE** (AM = 0.75)

Behaves like **MIN



Behaves like **MAX

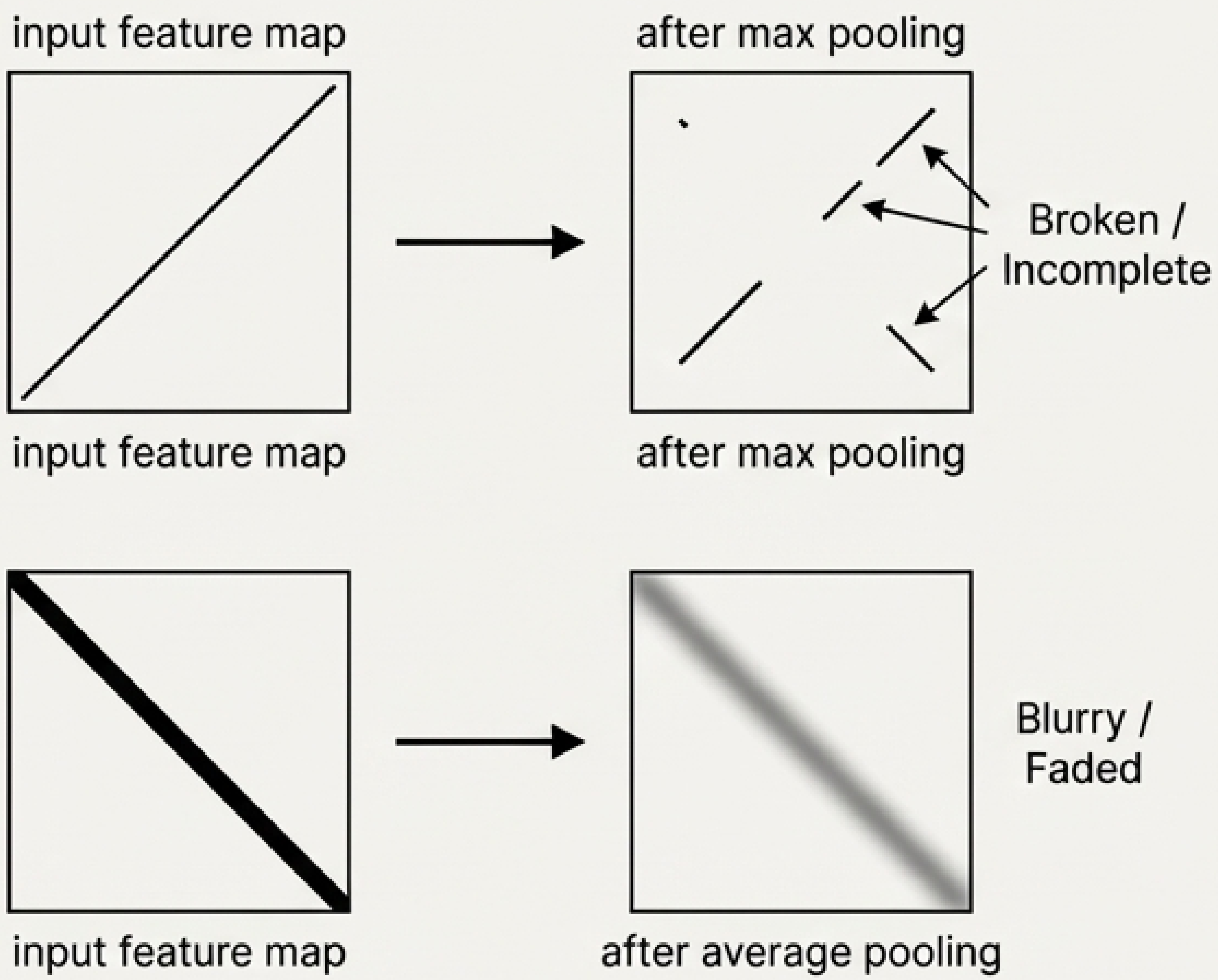
Behaves like **AVERAGE**

Behaves like **MIN

These surfaces show the output of GLM (top) and GPM (bottom) for the input vector (-2, -1, 0, 6). By learning to navigate this surface via the trainable parameters (α, β or γ, δ), the network can select the perfect functional behavior for its needs.

Application 1: Generalized Pooling Adapts to the Signal

❌ Fixed Pooling Fails



input feature map

after max pooling

Broken / Incomplete

input feature map

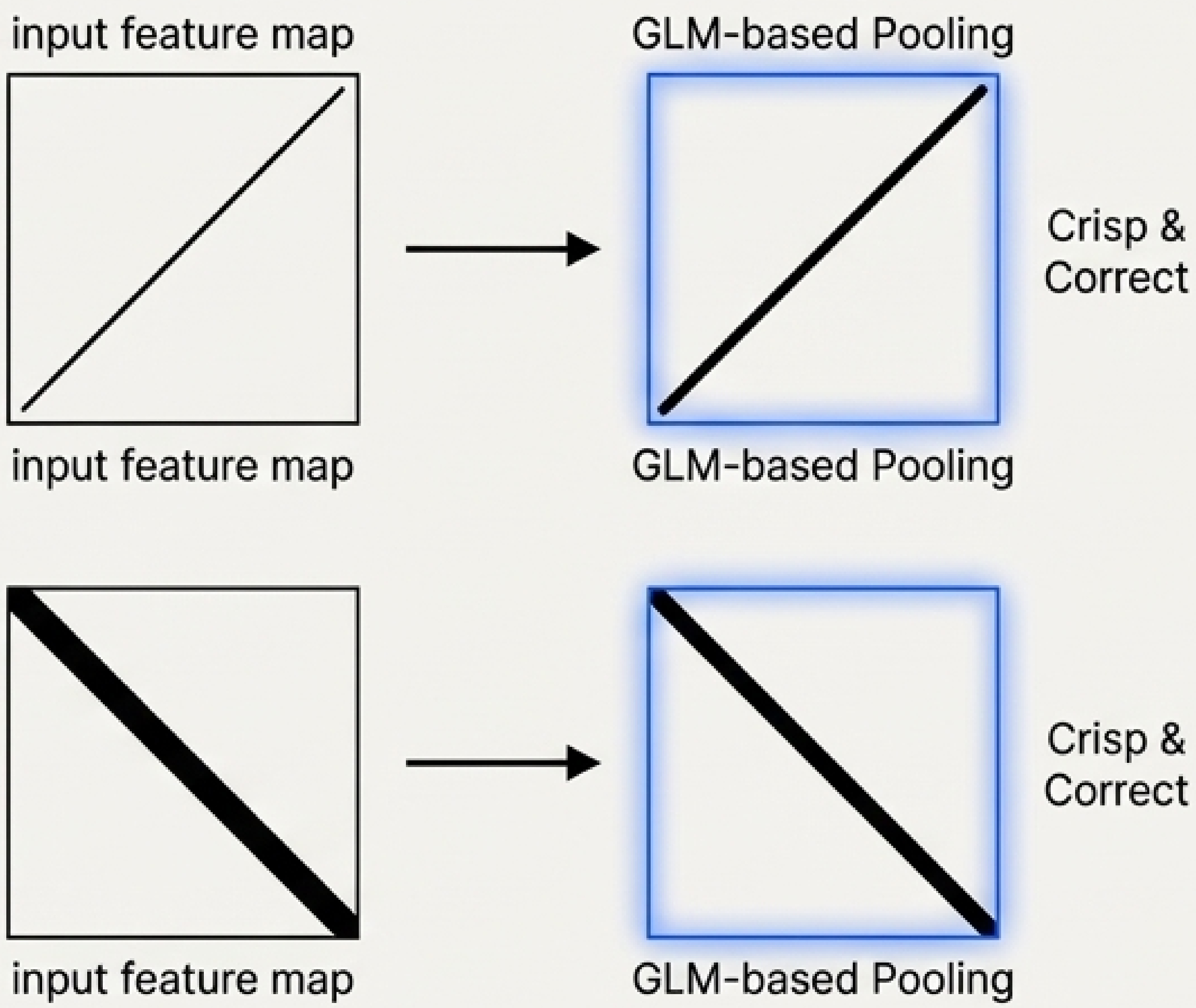
after max pooling

Blurry / Faded

input feature map

after average pooling

✅ Generalized Pooling Succeeds



input feature map

GLM-based Pooling

Crisp & Correct

input feature map

GLM-based Pooling

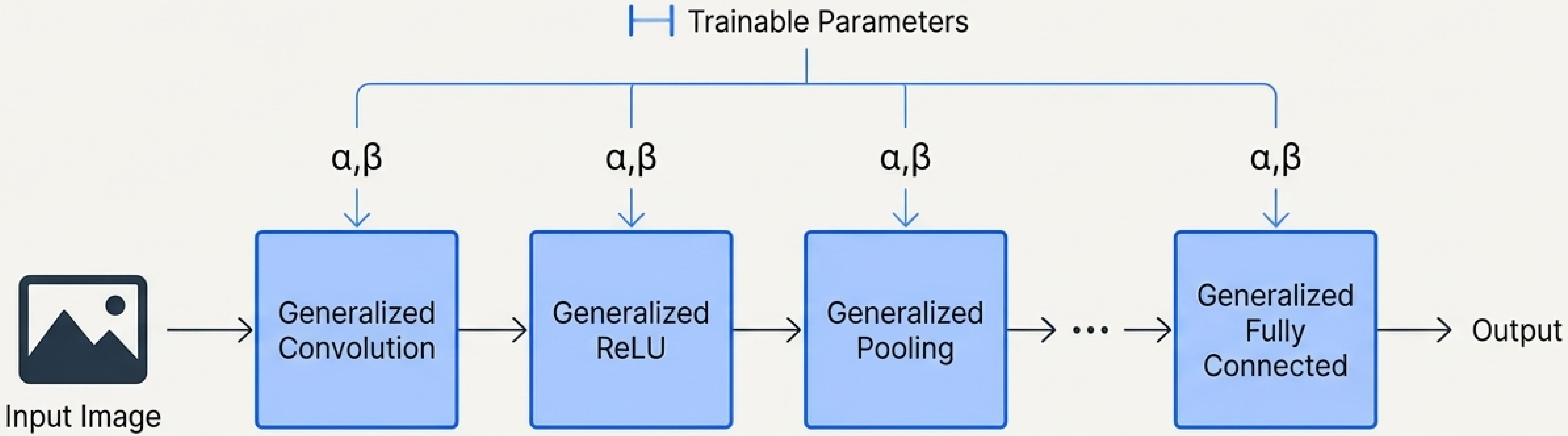
Crisp & Correct

input feature map

GLM-based Pooling

Unlike fixed methods, Generalized Pooling is sensitive to the *distribution* of values within the filter. It can learn to ignore outliers, average when needed, or select the strongest signal, leading to more robust feature extraction.

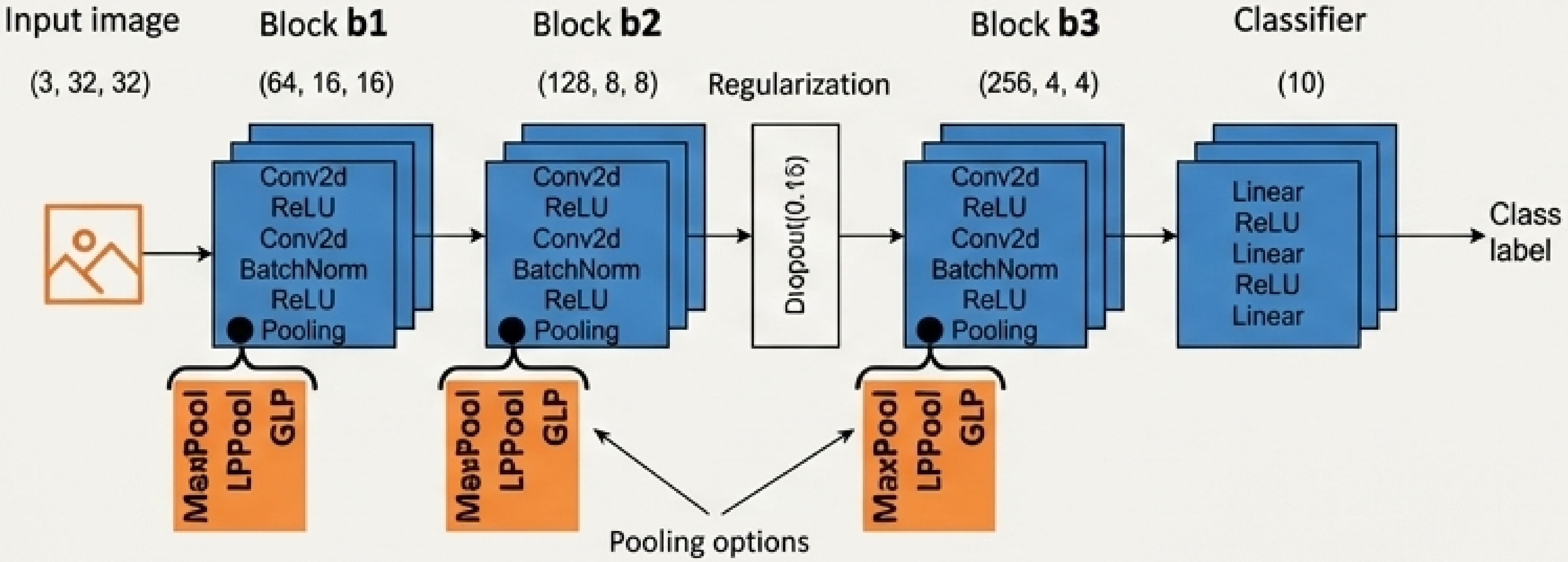
Upgrading the Entire CNN Stack with Learnable Components



The principle of mathematical flexibility can be applied to every core component. This creates a 'Hyper-Flexible' architecture where not just weights, but the fundamental operations themselves are learned and optimized end-to-end.

The Proving Ground: Testing Flexibility on CIFAR-10

Dataset & Architecture



Dataset: CIFAR-10 (60,000 32x32 color images in 10 classes).

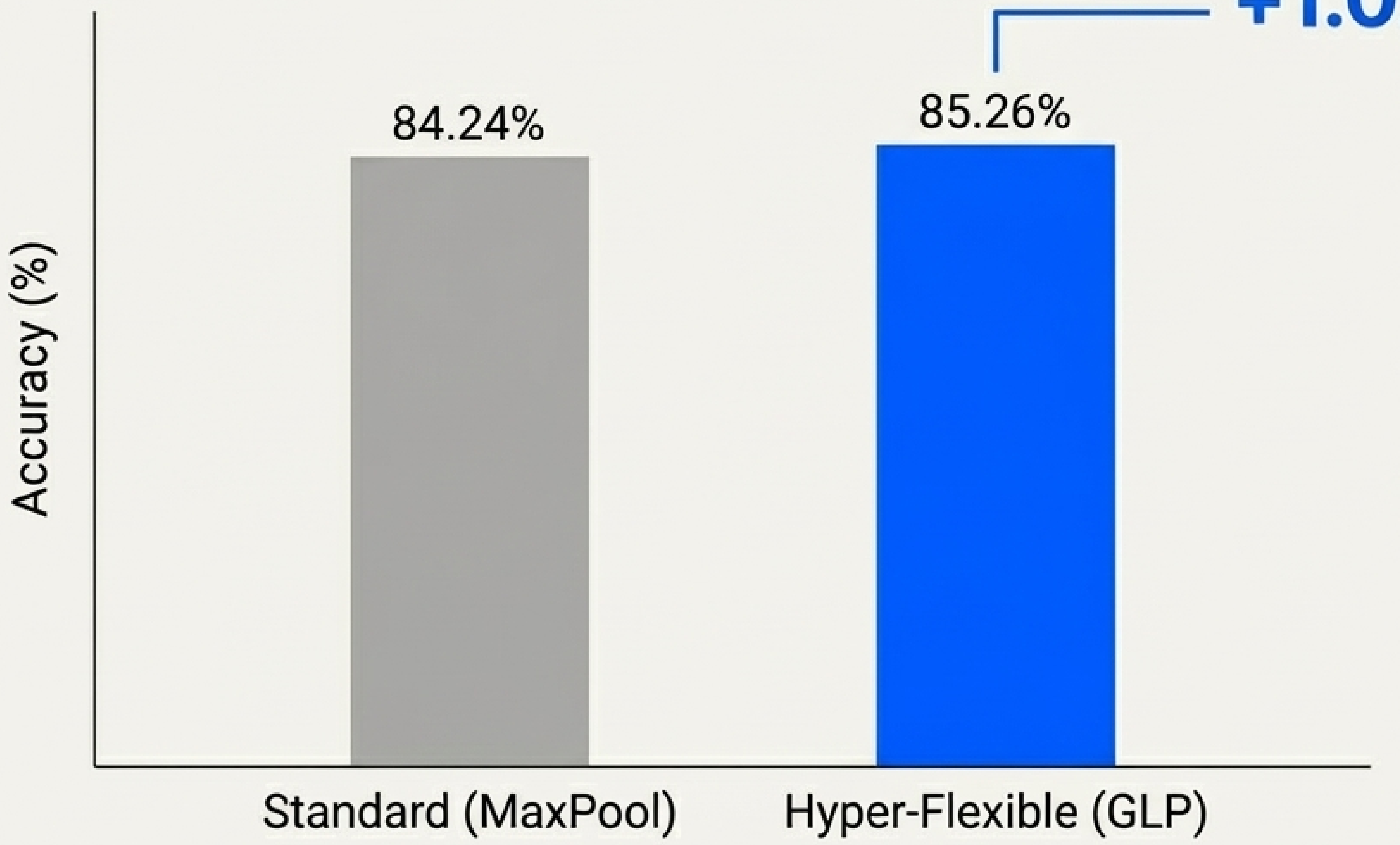
The Competitors

- **Standard CNN:** Uses fixed MaxPool.
- **Flexible CNN:** Uses LPPool (**Power Mean** with 1 trainable parameter).
- **Hyper-Flexible CNN:** Uses **Generalized Lehmer Pooling (GLP)** with 2 trainable parameters.

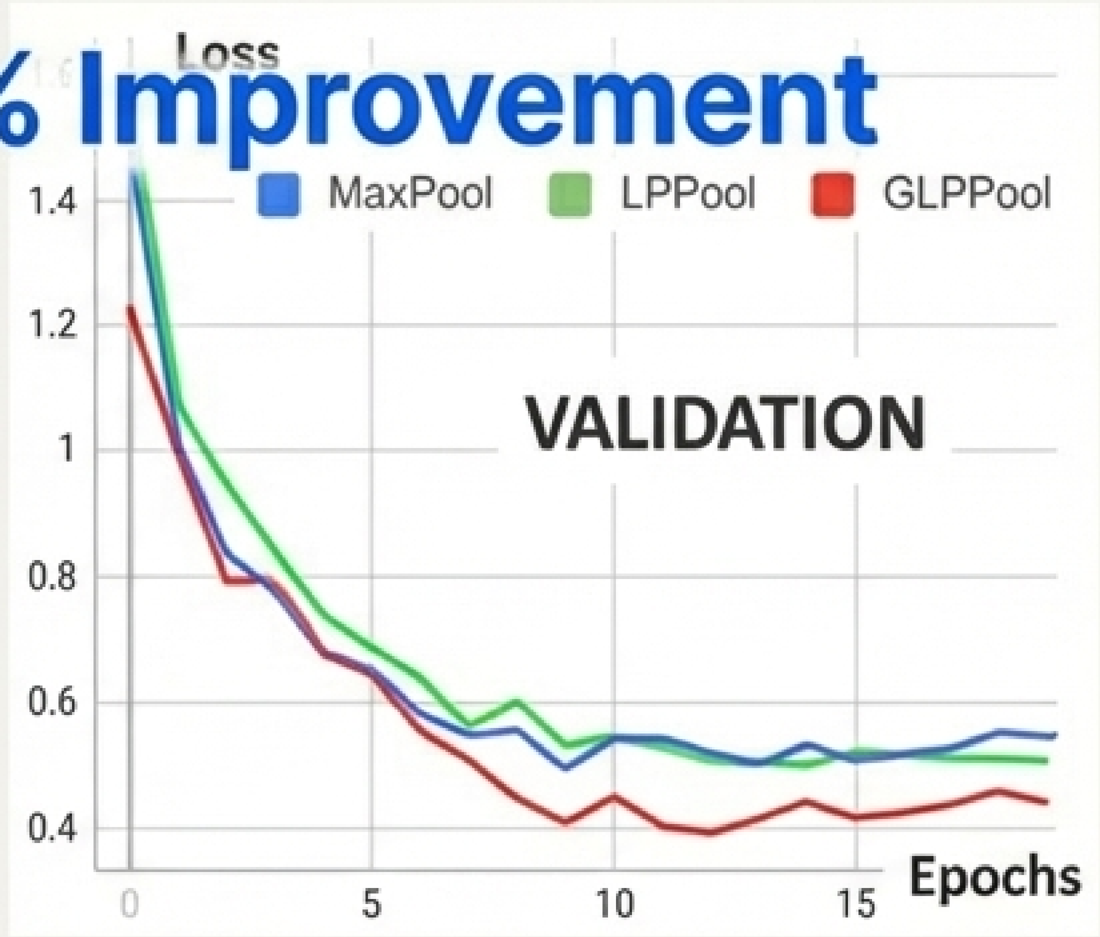
We compare these models to measure the impact of flexibility on both standard classification accuracy and robustness against targeted adversarial attacks.

Result 1: Flexibility Delivers a Clear Accuracy Advantage

Test Accuracy on CIFAR-10

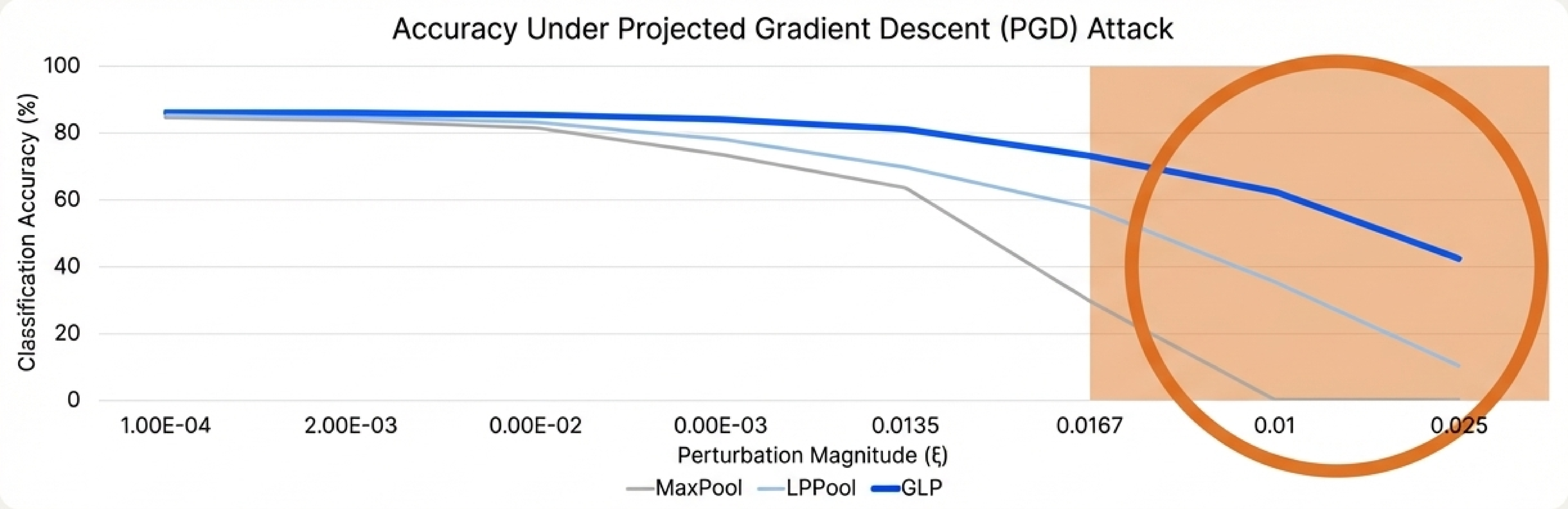


+1.02% Improvement



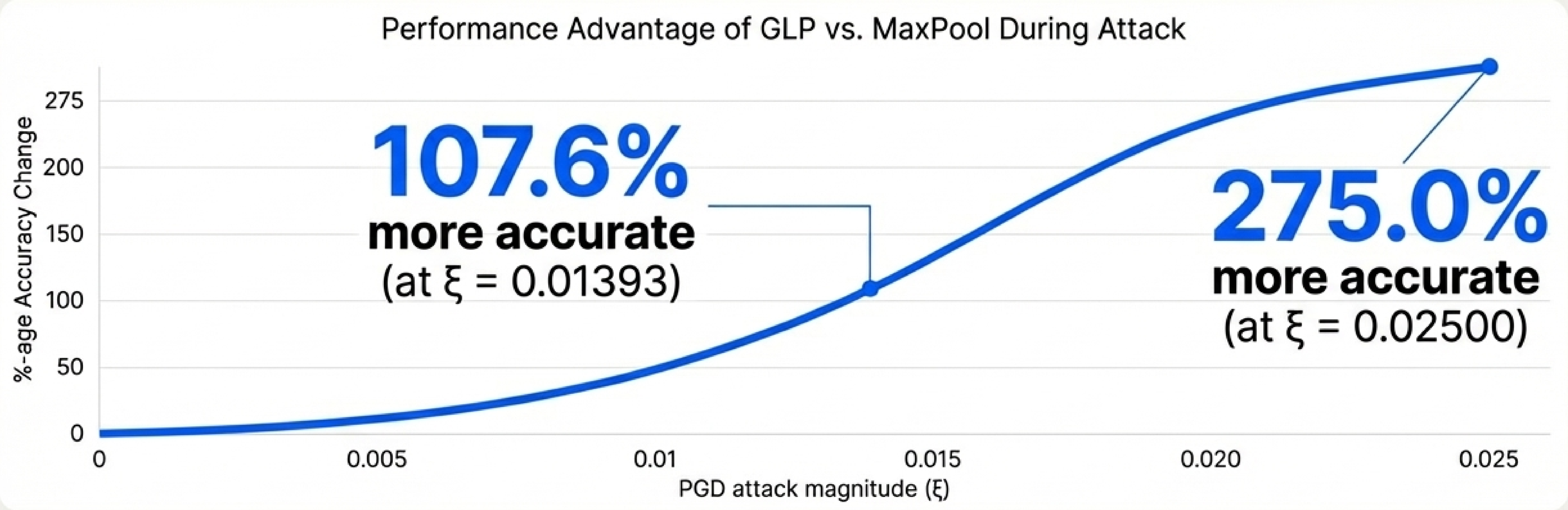
The validation loss for the GLP-based model (red) is consistently lower, indicating more stable and effective learning throughout training.

Result 2: Where Flexibility Truly Shines—Resisting Adversarial Attacks



As the attack intensity increases, the performance gap widens dramatically. While the standard model becomes unusable, the hyper-flexible model maintains significantly higher accuracy, demonstrating inherent resilience.

The Flexibility Dividend: Advantage Grows Exponentially Under Pressure



107.6% **more accurate** (at ξ = 0.01393)

275.0% **more accurate** (at ξ = 0.02500)

The more hostile the digital environment, the greater the return on mathematical flexibility. This is a fundamental principle for building truly resilient AI.

The Principle of Flexibility Extends Across the Architecture

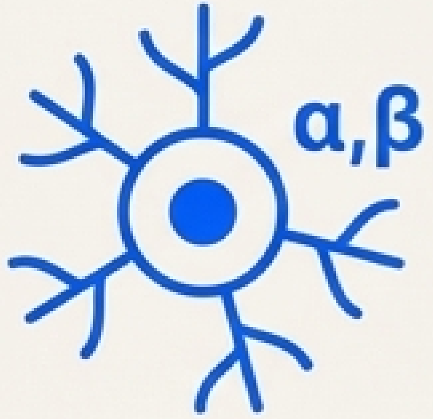
Generalized Convolution (GLC)

Replacing standard convolution with its flexible counterpart also boosts performance, with the greatest impact seen in the network’s deeper layers.

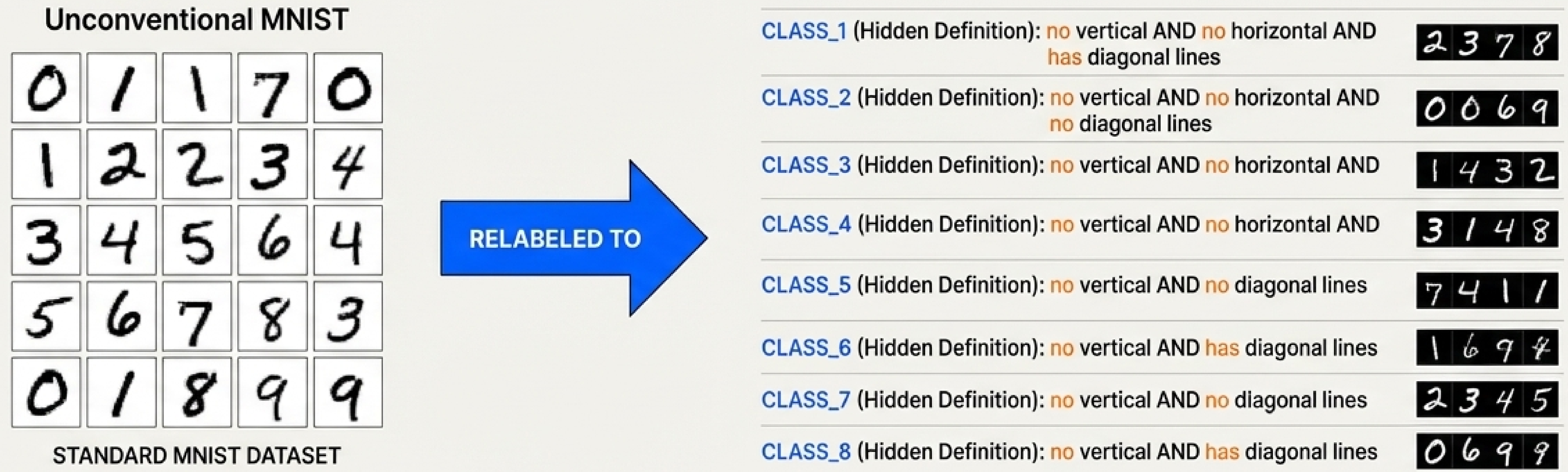
Standard CNN:	84.24%
CNN with GLC in block 3:	84.42%
CNN with GLC in all blocks:	84.88%

Generalized Neurons (GLN)

Even the basic neuron can be generalized. A Multi-Layer Perceptron built with GL-Neurons and SoftMax activation achieved **93.3%** accuracy on the Iris dataset, significantly outperforming a standard MLP’s 86.7%.



Beyond Pattern Matching: Tackling Unconventional Problems



Unconventional MNIST

STANDARD MNIST DATASET

RELABELED TO

CLASS_1 (Hidden Definition): no vertical AND no horizontal AND has diagonal lines

CLASS_2 (Hidden Definition): no vertical AND no horizontal AND no diagonal lines

CLASS_3 (Hidden Definition): no vertical AND no horizontal AND

CLASS_4 (Hidden Definition): no vertical AND no horizontal AND

CLASS_5 (Hidden Definition): no vertical AND no diagonal lines

CLASS_6 (Hidden Definition): no vertical AND has diagonal lines

CLASS_7 (Hidden Definition): no vertical AND no diagonal lines

CLASS_8 (Hidden Definition): no vertical AND has diagonal lines

Could a standard CNN, built on fixed MAX and SUM operations, learn to classify based on abstract, hidden logic like the absence of a feature? Its architecture is optimized for presence, not absence.

A hyper-flexible CNN is uniquely equipped for such challenges. It could learn to behave like MIN pooling to detect the absence of features or discover novel convolution behaviors to match abstract rules—something impossible for a rigid architecture.

From Fixed Math to Learned Capabilities

1

1. The Problem Defined

Standard CNNs are mathematically rigid, making them brittle and highly vulnerable to adversarial attacks.

2

2. A New Paradigm

We introduce Generalized Lehmer/Power Means (GLM/GPM)—parameterized functions that allow CNN components (Pooling, Convolution, Neurons) to *learn* their own optimal mathematical behavior.

3

3. The Proof Delivered

Hyper-flexible architectures demonstrate a modest accuracy gain on clean data but a **massive, compounding advantage in adversarial robustness**, proving the fundamental value of learned flexibility.

Explore the Future of Flexible AI

Future Directions

- Develop specialized training procedures to unlock the full potential of hyper-flexible models.
- Test performance on more complex datasets with non-obvious, hidden labeling logic.
- Apply generalized means to the backpropagation process itself for more adaptive learning.

Access the code, replicate the experiments, and build the next generation of robust neural networks.

`https://github.com/Adversarial-Intelligence-Group/flexnets`