Physical Rule-Guided Convolutional Neural Network

Kishor Datta Gupta
Clark Atlanta University
Atlanta, GA
kgupta@cau.edu

Marufa Kamal
BRAC University
Dhaka, Bangladesh
marufa.kamal1@g.bracu.ac.bd

Rakib Hossain Rifat
Texas Tech University
Lubbock, TX 79409, USA
rrifat@ttu.edu

Mohd Ariful Haque
Clark Atlanta University
Atlanta, GA
mohdariful.haque@students.cau.edu

Roy George
Clark Atlanta University
Atlanta, GA
rgeorge@cau.edu

Abstract

This paper presents a Physical-Guided Convolutional Neural Network (PGCNN) framework that incorporates dynamic, trainable, and automated LLM-generated, widely recognized rules integrated into the model as custom layers to address challenges like limited data and low confidence scores. The PGCNN is evaluated on multiple datasets, demonstrating superior performance compared to a baseline CNN model. Key improvements include a significant reduction in false positives and enhanced confidence scores for true detection. The results highlight the potential of PGCNNs to improve CNN performance for broader application areas.

1. Introduction

Deep learning models frequently struggle with object recognition because they rely on weight adjustments from training data,unlike humans who use shape, color, spatial cues, and context. For example, a human may use contextual awareness to recognize an object in water as a ship or debris, whereas a neural network may misclassify it in unrelated categories with high confidence. This demonstrates a gap in contextual comprehension. To address this, our approach incorporates context-aware weights generated by LLMs, allowing for adaptation across contexts. We believe that this can be expanded by utilizing more LLM-derived rules based on physical features to improve recognition. Recent research has investigated this direction through Physics-Guided Neural Networks (PGNNs), which shift beyond solely data-driven models by embedding physical rules, domain restrictions, and common-sense reasoning into neural architectures [13–15]. These approaches improve interpretability and performance, particularly in scientific and practical settings. Incorporating physical restrictions into loss functions can improve training outcomes [6]. These ideas apply not only to images, but also to text and audio inputs [1, 3, 4], where linguistic structure and semantics inform decisions.

In this paper, we present the Physical-Guided Convolutional Neural Network (PGCNN), a novel architecture that augments classic convolutional networks with dynamic, trainable layers based on physical reasoning and supplemented with domain knowledge supplied by the Large Language Model (LLM). PGCNN, which is based on the Faster R-CNN architecture and a ResNet-50 backbone, is tested on multi-environment datasets that include vehicle classes from both land and water scenarios. The framework is designed as a modular wrapper that can be smoothly merged with various detection models like YOLO. PGCNN includes custom layers that perform three core functions: bounding box refinement, which eliminates redundant or nested detections within the same class; scene-aware filtering, which uses dominant scene features (e.g., landmass or watermass) to influence classification decisions; and size-based reasoning, which validates predictions using relative object sizes, with inter-class relationships generated dynamically by OpenAI's LLM. These improvements reduce false positives or innaccurate and boost confidence in correct secondary labels—for example, reducing implausible predictions like "table" in a road scene and boosting more likely ones like "car" or "bus." Although PGCNN's mean Average Precision (mAP) is similar to the baseline, it considerably increases inference quality, trustworthiness, and interpretability. Its modular form makes it easy to apply to new domains, allowing for the merging of symbolic reasoning and deep learning.

2. Related Work

In recent years, significant advancements have been made in Physics-Guided Neural Networks (PGNNs). [6] developed a PGDNN combining neural networks with finite element models, achieving over 80% accuracy in structural damage detection. [16] introduced a PGNN for Fourier ptychographic microscopy, outperforming ePIE in high-defocus and high-exposure conditions. [2] reviewed over 250 studies on physics-informed computer vision (PICV), identifying PICV as highly effective. [8] showed PINNs and physics-guided nnU-Net excel in blood flow estimation using Doppler and CFD simulations. Hybrid CNN-PGNN models for aero-engine sensor diagnostics were explored by [7], while [5] used PG-BNN with Bayesian computation to predict concrete column strength. Inspired by these studies, we propose a novel PGCNN framework to further explore these applications.

3. Methodology

We present the Physical-Guided Convolutional Neural Network (PGCNN), an enhanced CNN architecture that integrates physical reasoning and domain-specific rules into custom layers. These layers are designed to improve interpretability and reduce false positives. Figure 1 provides an overview of the PGCNN framework, combining a Faster R-CNN ResNet-50 backbone with LLM-informed rule-based modules.

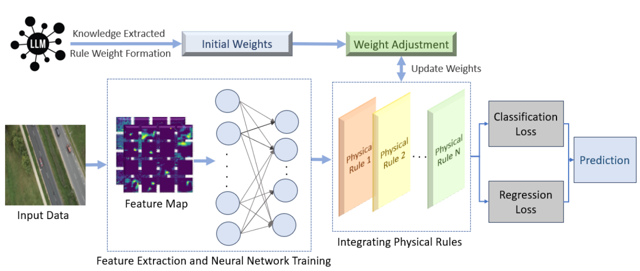


Figure 1. Overview of the PGCNN Framework

3.1. Shape-Based Detection Layer

This layer segments detected objects into geometric primitives (rectangles, triangles, etc.) and compares shape counts $S = \{s_1, s_2, \ldots, s_n\}$ against LLM-derived expected values $K = \{k_1, k_2, \ldots, k_n\}$. Confidence score C is adjusted using:

$$C = 1 - \frac{1}{1 + \exp\left(-\alpha \sum_{i=1}^{n} \left(\frac{s_i - k_i}{k_i}\right)^2\right)} \tag{1}$$

where α controls sensitivity to shape deviations.

3.2. Loss Function Integration

The custom layers feed context-aware feedback into the loss function to penalize incorrect detections and improve convergence. This enhances robustness and reliability during training.

3.3. Redundancy Elimination Layer

This layer filters redundant bounding boxes $B \in \mathbb{R}^{n \times 4}$ by comparing spatial overlap:

$$\text{Overlap}(B_i, B_j) = \frac{A(B_i \cap B_j)}{A(B_i)} \tag{2}$$

Bounding boxes are removed if one is fully enclosed or overlap exceeds a redundancy factor (RF).

3.4. Context-Aware Weight Adjustment Layer (CAWAL)

CAWAL adjusts logits during training based on contextual cues from scene predictions. If a scene contains a dominant feature (e.g., landmass), logit scores for related classes (e.g., cars) are increased by a factor α. The adjustment is conditionally applied when contextual tokens exceed a threshold Δt of scene labels. This layer enables the model to align predictions with environmental expectations.

3.5. Hybrid Weight Adjustment Layer (HWAD)

HWAD refines object logits using relative size constraints derived from an LLM-generated JSON rulebase. For each object pair, the model compares bounding box dimensions and updates class-wise logits when violations of known size relations occur. Posterior weights $P(T|E)$ are computed using Bayes' theorem:

$$P(T|E) = \frac{P(E|T) \cdot P(T)}{P(E)} \tag{3}$$

These values update rule confidence, blending LLM priors with empirical evidence using an update factor α.

Together, these physical rule layers enhance interpretability and reduce uncertainty, offering a robust mechanism for incorporating domain expertise into CNN-based object detection.

3.6. LLM Based Rule Generation

We generated the rules using OpenAI model **??** with weights in respect to the rule terms in between the classes. Figure 2 illustrates a snippet of the weights determined by the LLM for two classes. Initially, we investigated zero-shot prompting, in which the aim was to construct a JSON structure expressing a knowledge graph with vehicle types without providing explicit examples. We then supplied an example to instruct the model, which ensures consistent data creation and proper formatting.

```
{
  "@id": "/c/en/Bicycle",
  "@type": "node",
  "subClassOf": "Vehicle",
  "edges": [
    {
      "isSmallerThan": [
        { "@id": "/c/en/Bus", "weight": 1 },
        { "@id": "/c/en/Truck", "weight": 1 },
        { "@id": "/c/en/Car", "weight": 0.97 },
        { "@id": "/c/en/Motorcycle", "weight": 0.70 }
      ]
    }
  ]
},
{
  "@id": "/c/en/Motorcycle",
  "@type": "node",
  "subClassOf": "Vehicle",
  "edges": [
    {
      "isSmallerThan": [
        { "@id": "/c/en/Bus", "weight": 0.99 },
        { "@id": "/c/en/Truck", "weight": 0.95 },
        { "@id": "/c/en/Car", "weight": 0.94 }
      ],
      "isBiggerThan": { "@id": "/c/en/Bicycle", "weight": 0.70 }
    }
  ]
```

Figure 2. Example Snippet of Weights Generated by LLM For 'Motorcycle' and 'Bicycle' Class.

Few-Shot Prompt: Vehicle Size Graph Generation

Generate a JSON structure for vehicle classes ('Car', 'Bus', 'Truck', 'Bicycle', 'Motorcycle', 'Boat') using 'isSmallerThan' and 'isBiggerThan' edges with weights reflecting relative sizes.

Example for Bicycle:

```
{
  "@id": "/c/en/Bicycle",
  "@type": "node",
  "subClassOf": "Vehicle",
  "edges": [
    {
      "isSmallerThan": [
        { "@id": "/c/en/Bus", "weight": 1 },
        { "@id": "/c/en/Truck", "weight": 1 },
        { "@id": "/c/en/Car", "weight": 0.97 },
        { "@id": "/c/en/Motorcycle", "weight": 0.70 }
      ]
    }
  ]
}
```

Figure 3. Compact few-shot prompt for generating vehicle size graphs using edge-weighted comparisons.

4. Experimental Results and Discussion

4.1. Datasets

We conducted experiments using three datasets. The **Cars From Drone Dataset (CDD)** [12] includes 463 aerial images with five land vehicle classes. The **Drone Vehicle Dataset (DVD)** [11] consists of 17,927 images with five vehicle categories. Additionally, we curated a new **Multi-Environmental Vehicle Dataset (MEVD)**, combining the CDD classes with 177 annotated boat images sourced and converted into MS-COCO format [9].

4.2. Setup and Design

We use Faster R-CNN with ResNet-50 (pre-trained on ImageNet) and integrate three custom rule-based layers: redundancy elimination, scene-aware context adjustment (CAWAL), and hybrid weight adjustment (HWAD). A U-Net model [10] was used for scene segmentation.

4.3. Results

Table 1 summarizes mAP and IoU results. On the CDD dataset, PGCNN outperforms the baseline in mAP (0.450 vs. 0.420) and shows consistent IoU improvement at all thresholds. While DVD mAP is lower for PGCNN (0.221 vs. 0.325), it achieves better IoUs, suggesting enhanced localization but needing further optimization. On MEVD, both models yield similar mAP and IoU, except for a slight boost at IoU 0.9.

Table 1. Performance of Baseline vs. PGCNN

Metric	CDD		DVD		MEVD	
	Base	PGCNN	Base	PGCNN	Base	PGCNN
mAP	0.420	**0.450**	**0.325**	0.221	0.218	**0.218**
IoU@0.5	0.839	**0.851**	0.804	**0.813**	0.758	**0.758**
IoU@0.75	0.881	**0.903**	0.856	**0.858**	0.869	**0.869**
IoU@0.9	0.920	**0.926**	0.926	**0.927**	0.926	**0.929**

Figure 4 shows training loss on DVD. PGCNN consistently converges faster, starting at 0.34 and ending at 0.14, outperforming the baseline (0.49 to 0.21).

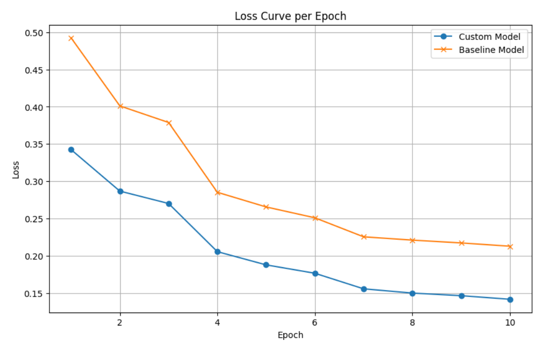


Figure 4. Training Loss Curve (DVD Dataset)

4.4. False Positive Mitigation

Bounding Box Reduction: As shown in Table 2, PGCNN significantly reduces redundant detections, achieving a 37.88% reduction in CDD and over 34% in MEVD.

Table 2. Bounding Box Reductions

Dataset	Baseline	PGCNN	Reduction (%)
CDD	598	451	37.88
MEVD	909	726	34.21
DVD	6192	5548	10.56

Mislabeled Box Reduction: Table 3 shows PGCNN reduces water vehicle false positives by 74.55% and land vehicle FPs by 39.01%, based on 78 images from MEVD.

Table 3. False Positive Reductions in MEVD

Class	Baseline	PGCNN	Reduction (%)
Water	110	28	74.55
Land	182	111	39.01

4.5. Confidence Score Optimization

The PGCNN model improved both high-confidence true detections and lowered confidence of false predictions. Table 4 summarizes these changes on MEVD (78 samples). Over 66% of correct predictions had increased confidence, and 71.9% of false positives saw reduced scores, reflecting improved contextual understanding.

Table 4. Change in Confidence Scores Before and After Rule Application on MEVD Dataset. "↑" indicates increased score confidence, while "↓" indicates reduced confidence, typically due to penalized rule violations.

Metric	Score↑	Score ↓
No. of Samples	395	430
Percentage (%)	66.05	71.91

Figures 5 and 6 illustrate sample predictions. PGCNN suppresses redundant bounding boxes and boosts confidence in context-aware predictions using physical constraints, demonstrating its capability to reduce model ambiguity and enhance interpretability.

Table 4 shows that PGCNN increases prediction reliability by raising confidence in right labels while suppressing plausible but wrong alternatives. This may not always increase precision in the classical sense (true positives), but it does reduce misunderstanding and ambiguity, which is critical in real-world deployment. Consider a model that forecasts an object as 85% boat, 70% table, and 60% ship. This prediction is less reliable than those for boats (85%), ships (65%), and tables (30%).

The second scenario demonstrates more consistent and physically realistic thinking, favoring useful categories while pushing down implausible ones. This type of "semantic disambiguation" is not captured by mAP but can be observed in qualitative results and human judgment.

Figures 5 and 6 show this impact. In Figure 6, the baseline awards 0.96 to a truck class among multiple possibilities, but the PGCNN model adjusts this to a more calibrated 0.76, delivering higher scores only when there's more contextual and visual alignment. This adjustment improves interpretability, prevents overconfidence in wrong options, and increases user trust in model results.

5. Conclusion and Future Works

Our primary focus was on integrating physical-attribute-based principles into CNN architecture to enhance the model's understanding beyond raw data features for the objects, rather than optimizing data selection strategies. We acknowledge that our method does not result in significant accuracy improvements, our target was to improving the confidence scores of each objets identified in a frame. Our model, which incorporates custom layers guided by LLM, is designed to be adaptable to any base model (e.g., RCNN, YOLO) without negatively affecting their performance through overfitting or underfitting. Our primary contribution lies in integrating physical attributes alongside raw data, enhancing the model's understanding. We have found that LLM can generate effective weight graphs to automate these custom layers, which can be retrained during the base model's training process. In the evaluation section, we recognize that measurements such as mAP are insensitive to such enhancements. As a result, in future research, we intend to include auxiliary metrics that evaluate "confidence concentration" or "prediction entropy reduction" across competing labels, as well as human evaluation tests for perceptual alignment and trust calibration.

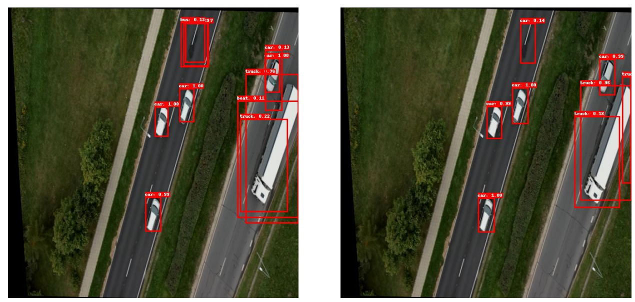
This pioneering approach not only improves the model's precision and reliability but also represents the first instance of incorporating rule-based modifications into a CNN network, paving the way for more dependable applications across diverse domains such as speech, robotics, healthcare, and natural language processing. Specially with areas where more reliable models are ought to be implemented.

Figure 5. Inference of Reduced Confidence Scores in FP

(a) Baseline model predicts two cars: one with high confidence and another redundant but accurate with a confidence score of 0.12.

(b) Custom model removes the redundant car Boundary box from the image.

Figure 6. Demonstration of Updated Confidence Scores



(a) Baseline model predicts truck with highest 0.76 score and more redundant boxes.

(b) Custom model detects truck with 0.96 score and the cars with either 1 or 0.99.

References

[1] William Apple, Lynn A Streeter, and Robert M Krauss. Effects of pitch and speech rate on personal attributions. *Journal of personality and social psychology*, 37(5):715, 1979. 1

[2] Chayan Banerjee, Kien Nguyen, Clinton Fookes, and George Karniadakis. Physics-informed computer vision: A review and perspectives. *arXiv preprint arXiv:2305.18035*, 2023. 2

[3] Douglas Biber. Text-linguistic approaches to register variation. *Register studies*, 1(1):42–75, 2019. 1

[4] Charles Clifton Jr and Susan A Duffy. Sentence and text comprehension: Roles of linguistic structure. *Annual Review of Psychology*, 52(1):167–196, 2001. 1

[5] Juan Fernández, Juan Chiachío, Manuel Chiachío, José Barros, and Matteo Corbetta. Physics-guided bayesian neural networks by abc-ss: Application to reinforced concrete columns. *Engineering Applications of Artificial Intelligence*, 119:105790, 2023. 2

[6] Zhou Huang, Xinfeng Yin, and Yang Liu. Physics-guided deep neural network for structural damage identification. *Ocean Engineering*, 260:112073, 2022. 1, 2

[7] Huihui Li, Linfeng Gou, Huacong Li, and Zhidan Liu. Physics-guided neural network model for aeroengine control system sensor fault diagnosis under dynamic conditions. *Aerospace*, 10(7):644, 2023. 2

[8] Hang Jung Ling, Salomé Bru, Julia Puig, Florian Vixège, Simon Mendez, Franck Nicoud, Pierre-Yves Courand, Olivier Bernard, and Damien Garcia. Physics-guided neural networks for intraventricular vector flow mapping. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024. 2

[9] Dataset Ninja. Visualization tools for ship detection from aerial images dataset. `https://datasetninja.com/aerial-ship-detection`, 2024. visited on 2024-08-02. 3

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 3

[11] Mranmay Shetty. Dronevehicle dataset. `https://universe.roboflow.com/mranmay-shetty/dronevehicle`, 2023. visited on 2024-08-13. 3

[12] VilniusTech University. Cars from drone dataset. `https://universe.roboflow.com/vilniustech-university/cars-from-drone-f5kny`, 2023. visited on 2024-08-02. 3

[13] Burhaneddin Yaman. *Self-Supervised Physics-Guided Deep Learning for Solving Inverse Problems in Imaging*. PhD thesis, University of Minnesota, 2022. 1

[14] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic resonance in medicine*, 84(6):3172–3191, 2020.

[15] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Ground-truth free multi-mask self-supervised physics-guided deep learning in highly accelerated mri. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1850–1854. IEEE, 2021. 1

[16] Yongbing Zhang, Yangzhe Liu, Xiu Li, Shaowei Jiang, Krishna Dixit, Xinfeng Zhang, and Xiangyang Ji. Pgnn: Physics-guided neural network for fourier ptychographic microscopy. *arXiv preprint arXiv:1909.08869*, 2019. 2