Improving Robust Image Classification Under Common Corruptions: A PDE-Regularized Variational Information Bottleneck Network

Gor Gharagyozyan * and Mariam Haroutunian *

Posted Date: 24 March 2026

doi: 10.20944/preprints202603.1823.v1

Keywords: image classification; corruption robustness; CIFAR-10-C; convolutional neural networks; partial differential equations; variational information bottleneck; information-theoretic learning; deep learning robustness

 doi:10.20944/preprints202603.1823.v1

Article

Improving Robust Image Classification Under Common Corruptions: A PDE-Regularized Variational Information Bottleneck Network

Gor Gharagyozyan * and Mariam Haroutunian *

Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia
* Correspondence: gor.gharagyozyan@edu.isec.am (G.G.); armar@sci.am (M.H.); Tel.: +374-55-599-565 (G.G.)

Abstract

Robust image classification remains challenging because deep Convolutional Neural Networks (CNNs) are highly sensitive to distribution shifts caused by common image corruptions such as noise, blur, and compression artifacts. To address this issue, this study investigates our proposed PDE-CNN-VIB architecture, which combines structural regularization derived from Partial Differential Equation (PDE) operators with information-theoretic feature compression based on the Variational Information Bottleneck (VIB) principle. The architecture was previously evaluated on the CIFAR-10 dataset and is further assessed here on both CIFAR-10 and its corrupted counterpart, CIFAR-10-C, using clean accuracy, negative log-likelihood (NLL), expected calibration error (ECE), and mean corruption accuracy (mCA). The model processes input images through a PDE-based regularization block and lightweight convolutional feature adaptation layers, followed by a VIB module that suppresses task-irrelevant information before the features are reconstructed and forwarded to a ResNet-18 classification backbone. The evaluation shows that the framework improves robustness under common corruptions, particularly for noise-related perturbations, while maintaining favorable clean-image performance and modest computational overhead. These findings indicate that combining PDE-based structural priors with VIB-driven feature compression is a promising approach for improving the reliability of CNNs under distribution shifts.

Keywords: image classification; corruption robustness; CIFAR-10-C; convolutional neural networks; partial differential equations; variational information bottleneck; information-theoretic learning; deep learning robustness

1. Introduction

Deep learning methods, particularly Convolutional Neural Networks (CNNs), have become the dominant paradigm for image classification and visual recognition tasks. During the past decade, CNN-based architectures have achieved remarkable performance in a wide range of applications, including object recognition, medical image analysis, and autonomous systems [1–3]. Despite these advances, modern CNN models remain vulnerable to distribution shifts between training and deployment environments. In particular, model performance often degrades significantly when images are affected by common corruptions such as noise, blur, weather effects, or compression artifacts [4,5]. This sensitivity represents a critical limitation for real-world applications, where input data rarely match the ideal conditions present in training datasets.

To systematically evaluate robustness under such distribution shifts, corruption benchmarks have been introduced. Among them, the CIFAR-10-C dataset has become a widely used benchmark to assess the robustness of the model under common image corruptions [4]. CIFAR-10-C extends the standard CIFAR-10 dataset [6] by introducing multiple types of corruption with varying severity levels, allowing researchers to quantify performance degradation under controlled perturbations. Metrics

 doi:10.20944/preprints202603.1823.v1

such as mean corruption accuracy (mCA) provide a comprehensive measure of robustness across corruption categories. Although many CNN architectures achieve high accuracy on clean datasets, their performance under corrupted conditions remains limited [7], motivating continued research on improving robustness and reliability of deep learning models under distribution shifts [8–10].

One promising direction involves integrating physics-inspired priors into deep learning models. In particular, convolutional filters derived from discretized Partial Differential Equations (PDEs) have been proposed to introduce spatial smoothing and structural regularization into neural networks. Such PDE-based filters can suppress high-frequency noise and enforce local spatial consistency in feature representations. Previous work introduced predefined PDE-based convolutional layers as structural components of CNN architectures, demonstrating improved feature stability and enhanced image recognition performance [11]. These approaches illustrate how physically motivated inductive biases can improve learning efficiency and robustness.

In parallel, information-theoretic learning frameworks have been investigated as mechanisms to control representation complexity and improve generalization in machine learning systems. The Information Bottleneck (IB) principle formulates learning as a trade-off between predictive accuracy and compression of the input representation [12–14]. Building upon this concept, the Variational Information Bottleneck (VIB) approach introduces a stochastic latent representation that encourages neural networks to retain only task-relevant information while suppressing nuisance variability [15,16]. Information-theoretic tools have also been widely studied as mechanisms for addressing various machine learning challenges, including generalization and robustness [17,18].

Our proposed PDE-CNN-VIB architecture combines PDE-based structural regularization with VIB-based feature compression within a unified image classification framework [19,20]. In previous work, the architecture was evaluated on the CIFAR-10 dataset, where improvements in generalization and calibration metrics such as the Expected Calibration Error (ECE) and Negative Log-Likelihood (NLL) were observed [20]. However, that evaluation focused primarily on clean image classification and did not assess robustness under distribution shifts caused by common image corruptions.

To address this limitation, the present study evaluates the proposed PDE-CNN-VIB model on the CIFAR-10-C benchmark, which contains multiple corruption types affecting image quality and structure. By extending the evaluation from the standard CIFAR-10 setting to the corrupted CIFAR-10-C setting, this paper provides a broader empirical picture of the architecture's behavior under common corruptions. The model combines PDE-based structural regularization with a convolutional feature extractor and a VIB module to improve robustness while maintaining competitive clean accuracy.

The main contributions of this work can be summarized as follows.

- We investigate the robustness of our proposed PDE-CNN-VIB architecture under common image corruptions using the CIFAR-10-C benchmark.
- We extend the evaluation of the proposed architecture beyond the standard CIFAR-10 setting to the corrupted CIFAR-10-C benchmark, providing a broader picture of its robustness behavior.
- We demonstrate improved corruption robustness compared to a baseline CNN, achieving higher mean corruption accuracy across multiple corruption types.
- We provide corruption-specific and category-wise analyses of robustness, highlighting substantial gains under high-frequency noise perturbations.
- We analyze the computational cost of the proposed model by comparing training time, inference latency, and parameter count with a baseline CNN.

The remainder of this paper is organized as follows. Section 2 describes the proposed architecture. Section 3 presents the datasets, corruption benchmark, training protocol, and evaluation metrics used in the study. Section 4 reports the experimental results on both clean and corrupted datasets. Section 5 discusses the implications of the findings, relates them to prior work, and outlines directions for future research. Finally, Section 6 concludes the article.

 doi:10.20944/preprints202603.1823.v1

2. Proposed Architecture

Consider our proposed architecture PDE-CNN-VIB, which combines PDE-based structural regularization, convolutional feature adaptation, and VIB compression within a unified classification framework [19,20]. The architecture was previously evaluated on the CIFAR-10 dataset and showed improved accuracy for clean images. In the present work, we assess the same architecture on the CIFAR-10-C benchmark to evaluate its robustness under corruption-induced distribution shifts and to provide a more comprehensive characterization of its behavior under corrupted inputs. Accordingly, the contribution of this paper lies in systematically assessing the robustness of the proposed framework under common corruptions through corruption-specific, category-wise, calibration, and computational-cost analyses.

As illustrated in Figure 1, the model consists of a custom PDE-CNN-VIB front-end that performs trainable diffusion-based regularization, lightweight convolutional feature adaptation, and stochastic compression through the VIB module. The resulting feature representation is then processed by a CIFAR-adapted ResNet-18 backbone and a linear classifier to produce the final prediction.

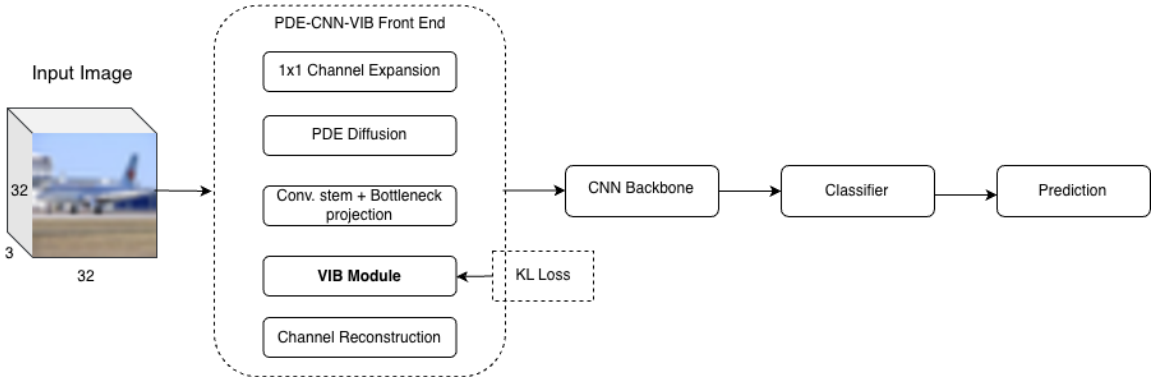

Figure 1. Overall architecture of the proposed PDE-CNN-VIB model.

2.1. PDE Regularization Layer

Physics-inspired priors have recently been explored as a way to improve the robustness and stability of deep learning models. In particular, PDE–based operators have long been used in image processing for tasks such as denoising, diffusion filtering, and edge-preserving smoothing [21,22]. These approaches rely on differential operators that model the evolution of image intensity values under diffusion-like processes and have proven effective for suppressing noise while preserving meaningful structural information.

Building on these ideas, PDE-inspired convolutional layers have been introduced as structural components of CNNs to incorporate physically motivated inductive biases into the learning process [11]. Such layers allow the network to perform diffusion-based regularization before hierarchical feature extraction, improving feature stability and reducing sensitivity to high-frequency perturbations.

In the proposed architecture, the PDE block combines a trainable channel-expansion layer with a fixed discrete Laplacian operator to introduce diffusion-based regularization into the feature maps. First, the input image is projected to a higher-dimensional feature space using a learnable 1×1 convolution that expands the number of channels. After this expansion, a depthwise PDE update is applied using a fixed Laplacian stencil. The diffusion strength is controlled by learnable nonnegative coefficients λ per-channel, allowing the model to adapt the amount of smoothing during training.

The resulting update follows a residual formulation corresponding to a discrete approximation of the classical heat diffusion equation [23],

$$u_{t+1} = u_t + \lambda \nabla^2 u_t, \tag{1}$$

where u_t denotes the feature map at iteration t, ∇^2 represents the discrete Laplacian operator, and λ is a learnable diffusion parameter applied independently to each channel. This formulation en-

 doi:10.20944/preprints202603.1823.v1

ables the model to suppress high-frequency noise while maintaining spatial structure in the feature representation.

The diffusion step can be repeated for a predefined number of iterations, allowing the network to progressively refine the feature representation through controlled diffusion. After the PDE regularization stage, the resulting feature maps are forwarded to the convolutional feature adaptation block, where they are refined and projected before stochastic compression by the VIB module.

2.2. Convolutional Feature Adaptation

After PDE-based regularization, the resulting feature maps are first processed by a lightweight convolutional front-end responsible for feature adaptation before stochastic compression. In the proposed implementation, this front-end consists of a small convolutional stem followed by a channel bottleneck projection. The convolutional stem refines the PDE-regularized representation while preserving spatial resolution, and the subsequent bottleneck projection reduces the channel dimension to produce a compact feature tensor for the VIB module.

These convolutional operations serve as the CNN component preceding the information bottleneck and motivate the designation PDE-CNN-VIB used in this work. Rather than applying the VIB module to the output of the full classification backbone, the bottleneck is introduced at this intermediate stage, where it can act directly on compact feature maps before deeper hierarchical processing.

2.3. VIB Module

To further improve the robustness and generalization of the learned feature representations, the proposed architecture incorporates a VIB module. The IB principle formulates learning as a trade-off between preserving information relevant for prediction and compressing irrelevant variations in the input representation [12]. By encouraging representations that retain only task-relevant information, the bottleneck framework can improve generalization and reduce sensitivity to noise and nuisance factors.

The VIB extends this concept by introducing a stochastic latent representation that approximates the optimal information-constrained encoding using variational inference techniques [15]. In this formulation, the representation of the deterministic feature produced by the neural network is transformed into a latent probabilistic variable characterized by mean μ and variance σ^2. Sampling from this distribution allows the model to learn compressed representations while maintaining predictive capability.

In the proposed PDE-CNN-VIB architecture, the VIB module is applied to bottlenecked feature maps produced by the lightweight convolutional front-end. First, the feature tensor is projected into a lower-dimensional channel space through a learnable 1×1 convolutional bottleneck layer. The VIB module then produces a stochastic representation z according to

$$z = \mu + \sigma \odot \epsilon, \tag{2}$$

where $\epsilon \sim \mathcal{N}(0, I)$ is a noise vector sampled from a standard Gaussian distribution and $\odot$ denotes element-wise multiplication. The resulting representation is then projected back to the original channel dimension and blended with the pre-VIB features through a residual connection before being passed to the ResNet-18 classification backbone.

Training is performed by optimizing a combined objective consisting of the standard classification loss and a Kullback–Leibler (KL) divergence [24] term that regularizes the latent distribution toward a unit Gaussian prior. The overall objective can be written as

$$\mathcal{L} = \mathcal{L}_{CE} + \beta D_{KL}(q(z|x) \parallel p(z)), \tag{3}$$

 doi:10.20944/preprints202603.1823.v1

where $\mathcal{L}_{CE}$ is the loss of the cross-entropy classification [15], $q(z|x)$ is the learned posterior distribution, $p(z)$ is the prior distribution, and β controls the strength of the information bottleneck. The coefficient β increases gradually during training using a linear warm-up schedule to stabilize optimization.

By introducing stochastic compression of intermediate features, the VIB module encourages the network to suppress nuisance information and focus on task-relevant features [20], which can improve robustness under distribution shifts such as image corruptions.

2.4. CNN Backbone

After stochastic compression and channel reconstruction, the resulting feature representation is processed by a CIFAR-adapted ResNet-18 backbone for final hierarchical feature extraction and classification. Residual networks have demonstrated strong performance and training stability by introducing identity shortcut connections that facilitate gradient propagation in deep architectures [3]. In this work, the ResNet-18 backbone operates on the reconstructed post-VIB feature maps and produces the final logits through a linear classification layer.

To accommodate the relatively small spatial resolution of CIFAR-10 images, we employ a CIFAR-adapted version of ResNet-18 that removes the initial 7×7 convolution and max-pooling layers typically used for large-scale datasets such as ImageNet. Instead, the network operates directly on the 32×32 input resolution, which has been shown to improve performance for small-image benchmarks.

The preceding section defines the architecture investigated in this study. The datasets, corruption benchmark, training configuration, and evaluation criteria used to assess the model on both clean and corrupted images are described in the next section.

3. Materials and Methods

This section describes the datasets, corruption benchmark, training protocol, and evaluation metrics used to assess the proposed PDE-CNN-VIB architecture. While previous work examined the architecture on clean CIFAR-10 data, the present experimental design is intended to provide a broader empirical picture by evaluating both standard classification performance and robustness under corruption-induced distribution shifts using CIFAR-10-C.

3.1. CIFAR-10 Dataset

The experiments in this study were conducted using the CIFAR-10 dataset, a widely used benchmark for evaluating image classification models in computer vision research [6]. In addition to its original role as a standard small-scale image classification benchmark, CIFAR-10 continues to serve as a common testbed in recent studies on data-efficient learning and dataset condensation [25,26]. The dataset contains 60,000 color images of size 32×32 pixels belonging to 10 object categories: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The dataset is split into 50,000 training images and 10,000 test images.

CIFAR-10 has been extensively used to study the performance and generalization properties of CNNs due to its balanced class distribution and manageable scale [1]. In this work, the dataset serves as the base dataset for training the proposed model under clean conditions, while robustness under corrupted conditions is evaluated using the CIFAR-10-C benchmark described in the following subsection.

During training, standard data augmentation techniques were applied to improve generalization. Each image was first padded with four pixels on each side and then a random 32×32 crop was extracted from the padded image, introducing small spatial translations while preserving the original resolution. In addition, random horizontal flipping was applied. After augmentation, the images were converted to tensors and normalized using the channel-wise mean $(0.4914, 0.4822, 0.4465)$ and standard deviation $(0.2470, 0.2435, 0.2616)$ computed from the training dataset. For evaluation on the clean test set, images were only converted to tensors and normalized using the same statistics.

 doi:10.20944/preprints202603.1823.v1

3.2. CIFAR-10-C Corruption Benchmark

To evaluate model robustness under distribution shift, we use the CIFAR-10-C corruption benchmark introduced by Hendrycks and Dietterich [4]. CIFAR-10-C extends the original CIFAR-10 dataset by introducing algorithmically generated corruptions that simulate common real-world image degradations. Such corruption benchmarks have become a standard tool for studying the robustness of deep learning models and have motivated the development of methods specifically designed to improve performance under corrupted inputs [27].

The dataset contains fifteen corruption types grouped into several categories, including noise, blur, weather effects, and digital distortions. Representative examples include Gaussian noise, shot noise, impulse noise, defocus blur, motion blur, snow, frost, fog, brightness variations, contrast changes, pixelation, and JPEG compression. Each corruption type is applied at five levels of severity, ranging from mild perturbations to strong distortions that significantly degrade image quality.

Figure 2 presents representative examples of CIFAR-10-C corruptions applied to the same input image, illustrating how different corruption types affect image appearance at a fixed severity level.

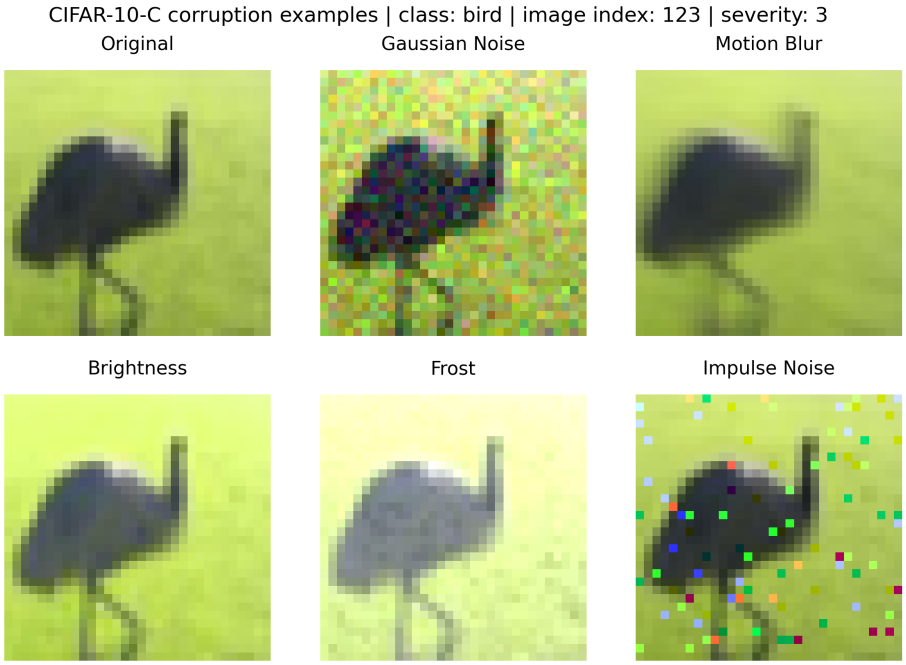

Figure 2. Representative examples of CIFAR-10-C corruptions applied to the same CIFAR-10 image at severity level 3. The figure illustrates how different corruption types, including Gaussian noise, motion blur, brightness, frost, and impulse noise, alter image structure and visual quality under controlled perturbations.

Following the standard evaluation protocol, robustness is quantified using the mean corruption accuracy (mCA), which measures the average classification accuracy across all corruption types. In this study, the results are reported at severity level 3, which represents a moderate level of degradation and is commonly used in robustness evaluations.

Using this benchmark, we compare the proposed PDE-CNN-VIB architecture with a baseline CNN trained under identical conditions. This setup allows us to quantify the extent to which the proposed method improves robustness under common image corruptions.

3.3. Training Protocol

All models were trained using the AdamW optimizer with an initial learning rate of 10^{-3}. Training was performed for 15 epochs with a batch size of 128. To improve generalization and stabilize training, label smoothing with a coefficient of 0.05 was applied to the cross-entropy loss function.

For models incorporating VIB, the overall objective consists of the standard classification loss combined with a KL divergence term that regularizes the latent representation. The weighting

 doi:10.20944/preprints202603.1823.v1

coefficient β controlling the strength of the bottleneck was gradually increased during training using a linear warm-up schedule. Specifically, β was initialized at a small value and increased linearly during the training process until reaching its maximum value.

The PDE regularization layer was applied before the convolutional feature adaptation and VIB stages, with the number of diffusion iterations fixed according to the architecture configuration. The ResNet-18 backbone then operated on the reconstructed post-VIB feature maps. All models were trained under identical data augmentation and preprocessing conditions to ensure a fair comparison between the baseline convolutional network and the proposed PDE-CNN-VIB architecture.

The experiments were conducted on a system equipped with an Apple M3 Pro processor and 36 GB of RAM running macOS Sonoma 14.4. The models were implemented in Python 3.11.9 using the PyTorch deep learning framework (version 2.10.0) together with Torchvision (version 0.25.0). Training and inference were accelerated using the Metal Performance Shaders (MPS) backend available on Apple Silicon hardware. In addition to classification performance metrics, computational efficiency was evaluated by measuring the number of trainable parameters, the training time per epoch, and the inference latency for both the baseline CNN and the proposed PDE-CNN-VIB architecture.

The exact architectural hyperparameters of the PDE-CNN-VIB model, including the channel expansion size, number of PDE diffusion iterations, VIB bottleneck dimension, and KL-weight warm-up configuration, are provided in the public implementation repository referenced in the Data Availability Statement. The repository also includes the complete training and evaluation scripts used to generate the reported results.

3.4. Evaluation Metrics

The performance of the models was evaluated using several complementary metrics that capture both the classification accuracy and the robustness properties. First, the standard top-1 classification accuracy was used to measure predictive performance on the clean CIFAR-10 test set.

To assess the calibration quality of the predicted probabilities, we computed the NLL and the ECE. NLL measures the quality of probabilistic predictions by evaluating the likelihood assigned to the correct class labels, while ECE quantifies the discrepancy between predicted confidence and empirical accuracy across probability bins.

Robustness under corrupted input conditions was evaluated using the mCA metric proposed by Hendrycks and Dietterich [4]. The mCA metric measures the average classification accuracy across multiple corruption types and provides a single robustness score that summarizes model performance under distribution shift. In this study, we report the mCA values computed on the set of corruption types included in the CIFAR-10-C benchmark at severity level 3.

In addition to accuracy-based metrics, computational efficiency was evaluated by comparing the number of trainable parameters, the average training time per epoch, and inference latency between the baseline CNN and the proposed PDE-CNN-VIB model. These measurements provide insight into the computational overhead introduced by the PDE regularization and VIB components.

4. Results

The experiments presented in this section evaluate the performance of the proposed PDE-CNN-VIB architecture and compare it with a baseline CNN trained under identical conditions. The evaluation considers both standard classification performance on clean data and robustness under distribution shifts introduced by image corruptions. Experiments are conducted using the CIFAR-10 dataset and the CIFAR-10-C corruption benchmark, and performance is measured using accuracy, NLL, ECE and mCA. The following subsections first present results on the clean dataset, followed by robustness evaluation under corrupted conditions and an analysis of the computational cost of the proposed method.

 doi:10.20944/preprints202603.1823.v1

4.1. Clean CIFAR-10 Performance

We first evaluate the classification performance of the proposed PDE-CNN-VIB architecture on the clean CIFAR-10 test set and compare it with a baseline CNN trained under identical conditions. The behavior of PDE-based architectures combined with VIB regularization under clean data conditions has previously been investigated, where improvements in generalization and calibration metrics were observed [20]. The present experiment aims to verify the effectiveness of the proposed architecture under the training configuration used in this study.

Table 1 summarizes the classification accuracy and calibration metrics for both models. In addition to standard accuracy, we report the NLL and ECE, which measure the quality of probabilistic predictions.

Table 1. Performance comparison on the clean CIFAR-10 test set.

Model	Accuracy (%)	NLL	ECE
Baseline CNN	80.60	0.584965	0.037420
PDE-CNN-VIB	**88.52**	**0.3684**	**0.0177**

The results indicate that the proposed architecture achieves substantially higher classification accuracy compared to the baseline CNN while also improving probabilistic calibration. In particular, the PDE-CNN-VIB model reduces both the NLL and the ECE, indicating more reliable confidence estimates. These findings are consistent with earlier observations that combining PDE-based structural regularization with VIB compression can improve feature stability and generalization.

4.2. Robustness on CIFAR-10-C

To evaluate robustness under distribution shifts caused by image degradations, we assess the performance of the proposed architecture using the CIFAR-10-C corruption benchmark [4]. This benchmark introduces fifteen types of image corruptions at multiple severity levels, allowing for systematic evaluation of how well a model maintains predictive performance under perturbed input conditions.

Table 2 reports the classification accuracy for each corruption type at severity level three for both the baseline CNN and the proposed PDE-CNN-VIB architecture. In addition to per-corruption accuracy, we report the mCA, which summarizes the overall robustness across all corruption types.

Table 2. Classification accuracy (%) on CIFAR-10-C corruption types (severity level 3).

Corruption Type	Baseline CNN	PDE-CNN-VIB
Gaussian Noise	25.79	39.62
Shot Noise	38.66	53.37
Impulse Noise	50.14	58.97
Defocus Blur	79.93	79.48
Glass Blur	47.82	57.48
Motion Blur	66.60	65.54
Zoom Blur	70.69	70.82
Snow	72.81	74.13
Frost	67.63	69.88
Fog	82.81	82.29
Brightness	86.14	86.49
Contrast	74.34	74.64
Elastic Transform	76.74	76.40
Pixelate	74.89	70.71
JPEG Compression	72.98	73.43
mCA	65.86	**68.88**

 doi:10.20944/preprints202603.1823.v1

The results demonstrate that the proposed PDE-CNN-VIB architecture consistently improves robustness across several corruption types. In particular, substantial improvements are observed for noise-related corruptions, where the proposed model significantly outperforms the baseline CNN. For example, under Gaussian noise, the classification accuracy increases from 25.79% to 39.62%, corresponding to an improvement of more than thirteen percentage points.

Overall, the PDE-CNN-VIB model achieves a higher mCA compared to the baseline CNN, indicating improved robustness under distribution shifts caused by common image corruption. These findings suggest that the combination of PDE-based structural regularization and VIB compression helps the network suppress noise-related perturbations while preserving task-relevant features.

4.3. Corruption Category Analysis

To further analyze the robustness characteristics of the proposed architecture, we examine the model performance across the different corruption types included in the CIFAR-10-C benchmark. Figure 3 illustrates the classification accuracy for each corruption type at severity level three for both the baseline CNN and the proposed PDE-CNN-VIB architecture.

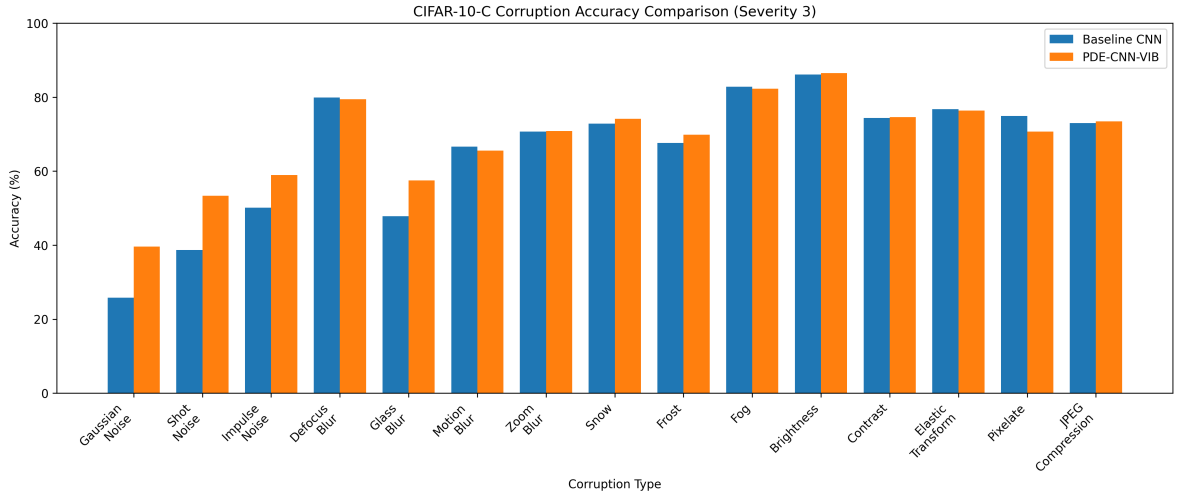

Figure 3. Classification accuracy for different CIFAR-10-C corruption types at severity level 3. The proposed PDE-CNN-VIB architecture demonstrates improved robustness particularly for noise-related corruptions compared with the baseline CNN.

The most significant improvements are observed for noise-related corruptions, including Gaussian noise, shot noise, and impulse noise. In particular, the classification accuracy under Gaussian noise increases from 25.79% for the baseline CNN to 39.62% for the PDE-CNN-VIB model, representing an improvement of more than thirteen percentage points. Similar improvements are observed for shot noise and impulse noise, indicating that the proposed architecture is more resilient to high-frequency perturbations.

This behavior can be explained by the design of the PDE regularization layer, which applies diffusion-based smoothing to the feature maps. Such diffusion processes naturally suppress high-frequency noise while preserving the overall spatial structure of the image. Consequently, the PDE layer helps stabilize the input representation before it is processed by the convolutional backbone.

For blur-related corruptions, such as defocus blur, glass blur, motion blur, and zoom blur, the differences between the two models are relatively small. These corruptions primarily remove spatial detail rather than introduce high-frequency noise, which explains why diffusion-based regularization has a more limited effect in these cases.

Weather-related corruptions, including snow, frost, and fog, show moderate improvements when using the proposed architecture. These perturbations combine noise-like patterns with changes in image contrast, which partially benefit from the smoothing behavior introduced by the PDE layer.

 doi:10.20944/preprints202603.1823.v1

Finally, digital distortions such as brightness changes, contrast variation, pixelation, and JPEG compression produce mixed results across the two models. Although improvements are modest in these cases, the overall robustness of the PDE-CNN-VIB architecture remains higher, as reflected by the increase in mean corruption accuracy reported in Table 2.

Figure 4 summarizes the average classification accuracy for each corruption category. The proposed PDE-CNN-VIB model demonstrates the largest improvements in the noise category, supporting the hypothesis that PDE-based diffusion helps suppress high-frequency perturbations.

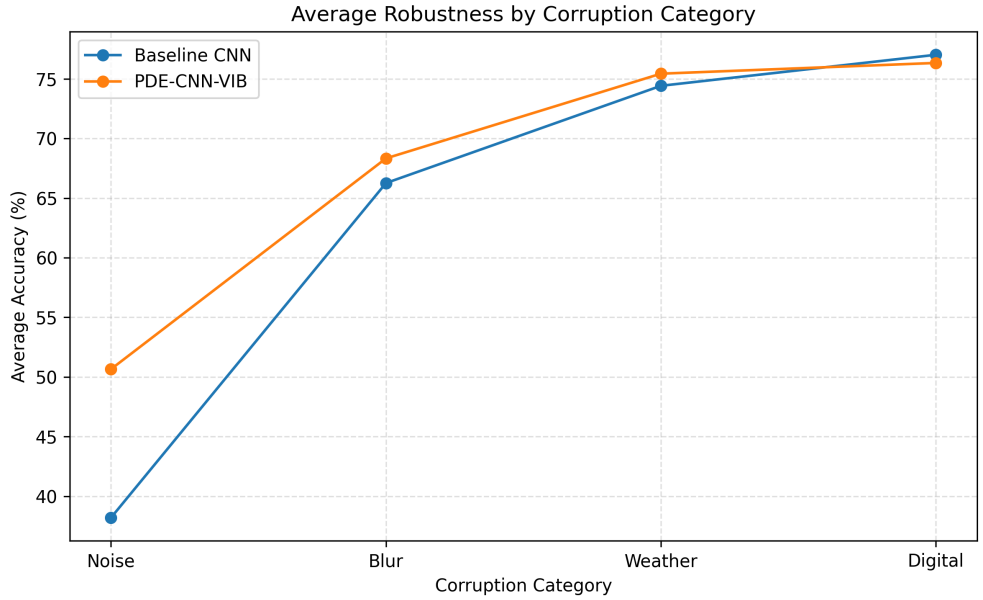

Figure 4. Average classification accuracy across corruption categories for the baseline CNN and the proposed PDE-CNN-VIB model. The proposed architecture shows the largest improvements for noise-related corruptions.

Overall, the results suggest that the combination of PDE-based structural regularization and VIB compression improves robustness primarily by reducing sensitivity to noise-related perturbations while preserving the discriminative information required for classification.

4.4. Computational Cost Analysis

In addition to classification performance, we evaluate the computational efficiency of the proposed architecture. Table 3 compares the number of trainable parameters, training time per epoch, and inference latency between the baseline CNN and the proposed PDE-CNN-VIB model.

Table 3. Comparison of computational cost between the baseline CNN and the proposed PDE-CNN-VIB architecture.

Model	Parameters (M)	Epoch Time (s)	Inference Time (ms/image)
Baseline CNN	11.17	108.77	1.463
PDE-CNN-VIB	11.20	122.12	1.516

The results show that the proposed model introduces only a small increase in computational cost. The number of trainable parameters increases by less than 0.3%, while the inference time increases by approximately 3.6%. Despite this modest computational overhead, the PDE-CNN-VIB architecture achieves significantly higher classification accuracy and robustness under corrupted conditions.

Figure 5 illustrates the trade-off between computational cost and robustness. The proposed architecture slightly increases inference time while providing a substantial improvement in mean corruption accuracy. This indicates that the additional computational cost is justified by the robustness gains achieved by the PDE-based regularization and VIB mechanisms.

 doi:10.20944/preprints202603.1823.v1

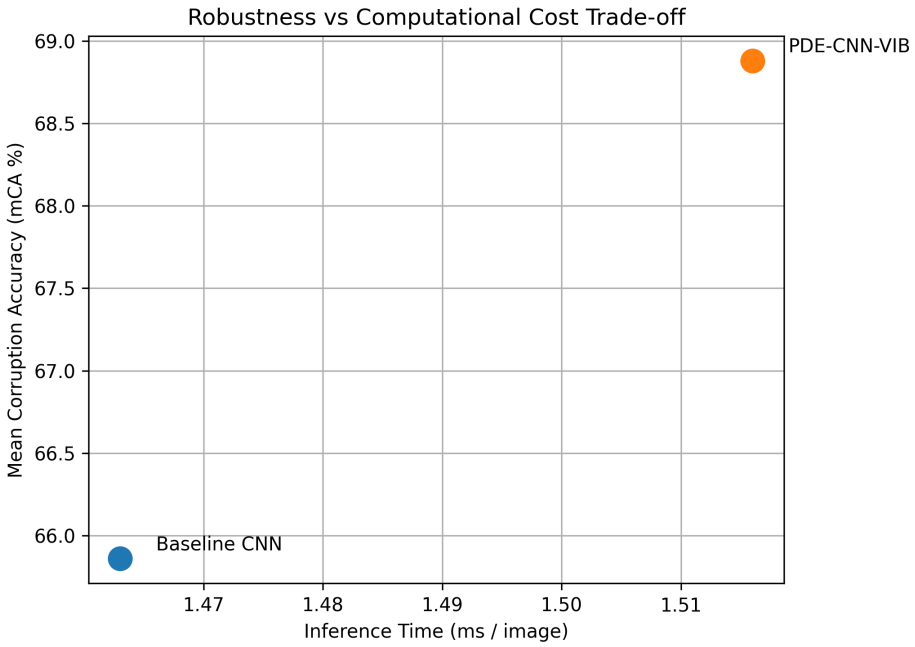

Figure 5. Trade-off between computational cost and robustness. The proposed PDE-CNN-VIB architecture introduces only a small increase in inference time while significantly improving mean corruption accuracy.

5. Discussion

The experimental results demonstrate that the proposed PDE-CNN-VIB architecture improves both classification performance and robustness under distribution shifts caused by common image corruptions. In particular, the model achieves higher mean corruption accuracy compared to the baseline CNN while introducing only a modest increase in computational cost. These findings suggest that combining physics-inspired structural priors with information-theoretic regularization can enhance the stability of learned representations.

One of the most notable observations is the substantial improvement for noise-related corruptions. The proposed architecture significantly increases classification accuracy for Gaussian, shot, and impulse noise perturbations. This behavior can be explained by the diffusion-based regularization performed by the PDE layer. Diffusion processes are known to suppress high-frequency noise while preserving the underlying structure of the signal, a property widely exploited in classical image processing methods [21,22]. By incorporating a discretized PDE operator directly into the neural network architecture, the model inherits this noise-suppressing behavior, which leads to improved robustness against high-frequency perturbations.

In addition to the PDE regularization, the VIB component contributes to the stability of the learned feature representations. The Information Bottleneck principle encourages neural networks to retain only task-relevant information while compressing irrelevant variations in the input representation [12]. The variational formulation introduced for deep neural networks enables this principle to be implemented efficiently during training [15]. Previous studies have shown that information bottleneck–based approaches can improve generalization and calibration in deep learning models [28]. In the present work, the VIB module further stabilizes the feature representations produced by the PDE-regularized backbone, reducing sensitivity to input perturbations.

Robustness to input perturbations has been widely studied in the deep learning literature, as neural networks often exhibit significant performance degradation when exposed to distribution shifts or adversarial perturbations [4,29]. Various approaches have been proposed to address robustness challenges, including uncertainty estimation, out-of-distribution detection, and advanced data augmentation strategies [27,30]. While many of these methods focus on improving training procedures, the

 doi:10.20944/preprints202603.1823.v1

results of this study indicate that incorporating structural priors directly into the network architecture provides an alternative pathway to improve the robustness.

Another important aspect of the proposed architecture is its relatively small computational overhead. Although the PDE-CNN-VIB model introduces additional operations compared to a standard convolutional network, the increase in computational cost remains limited. The number of trainable parameters grows only slightly, and both training time and inference latency increase by a small margin. Considering the significant improvements in robustness achieved by the model, this trade-off between performance and computational cost can be considered favorable for practical applications.

Despite the promising results, several limitations of the present study should be acknowledged. First, the experiments were conducted on the CIFAR-10-C benchmark, which, although widely used, represents a relatively small-scale dataset with low-resolution images. Evaluating the proposed architecture on larger and more complex datasets such as ImageNet-C would provide further insight into its scalability and general applicability. Second, the current implementation focuses on a specific PDE formulation based on a discrete Laplacian operator. Exploring alternative PDE formulations or adaptive diffusion mechanisms may further improve robustness.

Future work may also investigate integrating PDE-based regularization with other robustness-enhancing strategies, such as advanced data augmentation techniques [31] or adversarial training methods [32]. Additionally, studying the interaction between information-theoretic compression and physics-inspired priors in deeper architectures may reveal further improvements in robustness and generalization.

Overall, the results indicate that the combination of PDE-based structural regularization and VIB compression offers a promising direction to improve robustness in deep image classification models. By incorporating both physical modeling principles and information-theoretic constraints into neural network architectures, it is possible to enhance stability under distribution shifts while maintaining competitive computational efficiency.

6. Conclusions

This work investigated the robustness of our proposed PDE-CNN-VIB architecture under common image corruptions. Following its earlier evaluation on the CIFAR-10 dataset, the present study examined the architecture on the CIFAR-10-C benchmark in order to provide a broader picture of its behavior under corruption-induced distribution shifts. The main contribution of this paper is the corruption-robustness, calibration, and computational-cost assessment of the proposed framework.

The proposed model was evaluated on CIFAR-10 and the CIFAR-10-C benchmark. Compared with a baseline CNN, PDE-CNN-VIB achieved improved clean-image performance, better calibration, and higher mean corruption accuracy. The largest robustness gains were observed for noise-related corruptions, indicating that the diffusion-based PDE component is particularly effective at suppressing high-frequency perturbations, while the VIB module helps preserve task-relevant information.

At the same time, these improvements were obtained with only a small increase in computational cost, showing a favorable trade-off between robustness and efficiency. Overall, the results suggest that the integration of PDE-based structural priors with VIB-driven representation learning is a promising direction for developing more reliable image classification models under distribution shifts.

Author Contributions: Conceptualization, G.G. and M.H.; Methodology, G.G.; Software, G.G.; Validation, G.G.; Formal Analysis, G.G.; Investigation, G.G.; Data Curation, G.G.; Writing—Original Draft Preparation, G.G.; Writing—Review & Editing, G.G. and M.H.; Visualization, G.G.; Supervision, M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

 doi:10.20944/preprints202603.1823.v1

Data Availability Statement: The datasets analyzed in this study are publicly available. The CIFAR-10 dataset is available from its original public release, and the CIFAR-10-C corruption benchmark is available from the repository provided by Hendrycks and Dietterich [4]. The implementation of the proposed PDE-CNN-VIB model, including training and evaluation scripts used to reproduce the results presented in this paper, is publicly available at: https://github.com/gor-gh/pde-cnn-vib-cifar10c.

Use of Artificial Intelligence: Generative artificial intelligence was used to assist with language refinement and structural editing of the manuscript. The authors reviewed, revised, and validated all generated text and take full responsibility for the content of the article. No AI tools were used to generate data, perform experiments, or produce the reported results.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
PDE	Partial Differential Equation
VIB	Variational Information Bottleneck
mCA	Mean Corruption Accuracy
NLL	Negative Log-Likelihood
ECE	Expected Calibration Error
KL	Kullback–Leibler Divergence
CIFAR	Canadian Institute for Advanced Research
MPS	Metal Performance Shaders

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
2. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems* **2012**, *10*, 1097–1105.
3. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, 770–778.
4. Hendrycks, D.; Dietterich, T. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, March 2019.
5. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.; Brendel, W. Generalisation in Humans and Deep Neural Networks. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, QC, Canada, December 2018, 7538–7550.
6. Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. Technical Report, University of Toronto, Toronto, ON, Canada, 2009.
7. Aghababaeyan, N. Towards Reliable Image Classification: A Systematic Robustness Analysis of CNN and Classical Models Under Natural Corruptions. Master's Thesis, Concordia University, Montreal, Canada, 2025.
8. Mitra, P.; Schwalbe, G.; Klein, N. Investigating Calibration and Corruption Robustness of Post-hoc Pruned Perception CNNs: An Image Classification Benchmark Study. arXiv preprint arXiv:2405.20876, 2024.
9. Tran, A.T.; Zeevi, T.; Payabvash, S. Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging. BioMedInformatics 2025, 5, 20. https://doi.org/10.3390/biomedinformatics5020020
10. Huang, Z. Trustworthy Machine Learning under Distribution Shifts. *arXiv* 2025, arXiv:2512.23524.
11. Sahakyan, V.; Melkonyan, V.; Gharagyozyan, G.; Avetisyan, A. Enhancing Image Recognition with Pre-Defined Convolutional Layers Based on PDEs. *Programming and Computer Software* **2023**, *49(3)*, 192–197.
12. Tishby, N.; Pereira, F.; Bialek, W. The Information Bottleneck Method. In Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 1999, 368–377.
13. Harremoës, P.; Tishby, N. The Information Bottleneck Revisited or How to Choose a Good Distortion Measure. 2007 IEEE International Symposium on Information Theory (ISIT), Nice, France, 2007, 566–570.

 doi:10.20944/preprints202603.1823.v1

14. Shwartz-Ziv, R.; Tishby, N. The Role of the Information Bottleneck in Representation Learning. International Conference on Learning Representations (ICLR) Workshop on Deep Learning Theory, Toulon, France, 2017.
15. Alemi, A.; Fischer, I.; Dillon, J.; Murphy, K. Deep Variational Information Bottleneck. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 2017.
16. Wang, C.; Du, S.; Zhang, Y. Self-Supervised Learning for High-Resolution Remote Sensing Images Change Detection With Variational Information Bottleneck. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2023**.
17. Haroutunian, M. E.; Gharagyozyan, G. A. Information Theory Tools and Techniques to Overcome Machine Learning Challenges. *Mathematical Problems of Computer Science* **2025**, *63*, 25–41.
18. Feder, M.; Urbanke, R.; Fogel, Y. Information-Theoretic Framework for Understanding Modern Machine Learning. *arXiv* 2025, arXiv:2506.07661.
19. Gharagyozyan, G. Improving CNN Generalization with PDE Preprocessing and the Variational Information Bottleneck. In Proceedings of the International Conference on Computer Science and Information Technologies (CSIT), Yerevan, Armenia, September 2025, 145–147.
20. Gharagyozyan, G. A PDE-Based Convolutional Neural Network with Variational Information Bottleneck: Experimental Evaluation and Generalization Analysis. *Mathematical Problems of Computer Science* **2025**, *64*, 37–46.
21. Perona, P.; Malik, J. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1990**, *12(7)*, 629–639.
22. Weickert, J. *Anisotropic Diffusion in Image Processing*; Teubner: Stuttgart, Germany, 1998.
23. Richtmyer, R. D.; Morton, K. W. *Difference Methods for Initial-Value Problems*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1967; pp. 1–30.
24. Cover, T. M.; Thomas, J. A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2006.
25. Mohanty, S.; Reddy, A.; Mopuri, K. R. DiRe: Diversity-promoting Regularization for Dataset Condensation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA, 2026.
26. Qi, D.; Li, J.; Dou, S.; Gao, J.; Wang, Y.; Zhao, B. Active Dataset Distillation via Dual-Space Informative Matching. *IEEE Transactions on Image Processing* **2026**.
27. Hendrycks, D.; Mu, N.; Cubuk, E.; Zoph, B.; Gilmer, J.; Lakshminarayanan, B. AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty. In Proceedings of the International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 2020.
28. Tishby, N.; Zaslavsky, N. Deep Learning and the Information Bottleneck Principle. In Proceedings of the 2015 IEEE Information Theory Workshop (ITW), Jerusalem, Israel, 2015, 1–5.
29. Goodfellow, I.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.
30. Hendrycks, D.; Gimpel, K. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 2017.
31. Cubuk, E.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, Q. AutoAugment: Learning Augmentation Strategies from Data. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, 113–123.
32. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards Deep Learning Models Resistant to Adversarial Attacks. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 2018.