ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

Ղարագյոզյան Գոռ Արթուրի

ԻՆՖՈՐՄԱՑԻԱՅԻ ՏԵՍՈՒԹՅԱՆ ԳՈՐԾԻՔԱԿԱԶՄԻ ՆԵՐԴՐՈՒՄԸ ԽՈՐՔԱՅԻՆ ՆԵՅՐՈՆԱՅԻՆ ՑԱՆՑԻ ՃԱՐՏԱՐԱՊԵՏՈՒԹՅՈՒՆՈՒՄ՝ ՊԱՏԿԵՐՆԵՐԻ ԴԱՍԱԿԱՐԳՄԱՆ ԱՐԴՅՈՒՆԱՎԵՏՈՒԹՅԱՆ ԲԱՐՁՐԱՑՄԱՆ ՆՊԱՏԱԿՈՎ

Ե.13.05 - «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի համար

ՍԵՂՄԱԳԻՐ

Երևան 2026

INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF THE NAS RA

Gharagyozyan Gor

INTEGRATION OF INFORMATION-THEORETIC TOOLS INTO DEEP NEURAL NETWORK ARCHITECTURE FOR IMPROVING IMAGE CLASSIFICATION PERFORMANCE

SYNOPSIS

of the dissertation for obtaining a Ph.D. degree in Technical Sciences on specialty 05.13.05 “Mathematical modelling, numerical methods and program complexes”

Yerevan 2026

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում:

Գիտական ղեկավար՝ ֆիզ. մաթ. գիտ. դոկտոր. Մ. Ե. Հարությունյան

Պաշտոնական ընդդիմախոսներ՝ տեխ. գիտ. դոկտոր Դ. Գ. Ասատրյան

տեխ. գիտ. թեկնածու Վ. Կ. Ավետիսյան

Առաջատար կազմակերպություն՝ Երևանի պետական համալսարան

Ատենախոսության պաշտպանությունը տեղի կունենա 2026թ. Հուլիսի 3-ին՝ ժամը 14:00-ին, ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող 037 «Ինֆորմատիկա» մասնագիտական խորհրդի նիստում հետևյալ հասցեով՝ Երևան, 0014, Պ. Սևակի 1:

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ գրադարանում:
Սեղմագիրն առաքված է 2026թ. Հունիսի 3-ին:

Մասնագիտական խորհրդի գիտական

քարտուղար ֆիզ. մաթ. գիտ. դոկտոր՝ Մ. Ե. Հարությունյան

The topic of the dissertation was approved at the Institute of Informatics and Automation Problems of NAS RA.

Scientific supervisor: Doctor of phys-math sciences M. E. Haroutunian

Ofiicial opponents: Doctor of technical sciences D. G. Asatryan
Candidate of technical sciences V. K. Avetisyan

Leading organization: Yerevan State University

The Defense will take place on the 3th of July 2026 at 14:00, at the Specialized Council 037 “Informatics” at the Institute of Informatics and Automation Problems of NAS RA.
Address: Yerevan, 0014, P. Sevak 1.

The Dissertation is available at the library of IIAP NAS RA.
The synopsis is delivered on 3th of June 2026.

Scientific Secretary of the Specialized Council, D.Ph.M.S. M. E. Haroutunian

General description of the work

Relevance of the topic

In the twenty-first century, the rapid development of digital technologies, Artificial Intelligence, and Machine Learning (ML) has significantly influenced the way visual information is processed, analyzed, and used in decision-making systems. Digital images are now widely used in scientific research, medical diagnosis, autonomous systems, remote sensing, surveillance, industrial inspection, and many other applied domains. In these areas, image classification systems are expected not only to recognize visual objects under ideal conditions, but also to remain reliable when the input data are affected by noise, blur, compression artifacts, illumination changes, weather effects, or other forms of degradation.

Deep Convolutional Neural Networks (CNN) have become the dominant approach for image classification due to their ability to learn hierarchical representations directly from image data. Modern CNN architectures can achieve high performance on standard benchmark datasets and have become the basis of many visual recognition systems. However, their strong performance under clean benchmark conditions does not guarantee reliable behavior in real-world environments. In practice, images may differ from the training data distribution because of sensor limitations, environmental changes, motion, compression, or corruption. Under such conditions, CNN classifiers may lose accuracy, become poorly calibrated, or produce overconfident incorrect predictions. Therefore, robust image classification remains an important and practically relevant problem.

The relevance of this problem is especially clear in application areas where incorrect visual decisions may lead to serious consequences. In autonomous driving and driver-assistance systems, camera images may be affected by rain, fog, snow, low illumination, or motion blur. In medical image analysis, diagnostic images may contain acquisition noise, low contrast, or scanner-dependent artifacts. In remote sensing and drone imagery, atmospheric effects, compression, and illumination changes may degrade visual quality. Similar problems occur in surveillance, traffic monitoring, industrial visual inspection, and edge-based vision systems. In all these cases, the reliability of image classification depends not only on clean-data accuracy, but also on robustness under non-ideal visual conditions.

The need for robustness has led to the development of corruption-oriented benchmarks such as CIFAR-10-C and CIFAR-100-C, where standard image datasets are extended with controlled corruptions. These benchmarks make it possible to measure how much model performance degrades under different types of image distortion. Such evaluation is important because it reveals weaknesses that are not visible when only clean test accuracy is reported. A model that performs well on clean CIFAR-10 or CIFAR-100 may still be sensitive to Gaussian noise, impulse noise, motion blur, contrast changes, pixelation, or JPEG compression. Thus, robustness-oriented evaluation has become a necessary part of studying the practical reliability of CNN-based classifiers.

At the same time, improving robustness requires more than simply increasing model size or training longer. Modern CNNs may learn representations that contain redundant, nuisance, or corruption-sensitive information. Information-Theoretic tools provide a principled way to analyze and control such representations. In particular, the Information Bottleneck (IB)[1] principle and its variational formulation encourage a model to preserve task-relevant information while compressing irrelevant variability. This

[1] Tishby N., Pereira F. C. and Bialek W., “The information bottleneck method”, arXiv:physics/0004057, 1999.

makes the Variational Information Bottleneck (VIB)[2] especially relevant for improving generalization, calibration, and stability of neural-network representations.

However, Information-Theoretic compression alone does not explicitly impose spatial structure on visual features. Image data have strong local spatial organization: neighboring pixels form edges, textures, contours, and object parts. Therefore, robustness in image classification also depends on the stability of early spatial feature extraction. Partial differential Equation (PDE)[3] based methods are relevant in this context because they provide mathematically interpretable mechanisms for smoothing, diffusion, and structural regularization. The work also develops PDE-based convolutional prior layers as front-end components and extends them into a lightweight diffusion-based trainable PDE module with learnable nonnegative channel-wise diffusion coefficients for improving low-level feature extraction in CNN-based image classification.

Thus, the relevance of the dissertation is determined by the need to develop CNN-based image classification models that are not only accurate on clean data, but also more stable, better calibrated, and more robust under corrupted visual conditions. The work addresses this need by combining two complementary principles: diffusion-based PDE regularization for stabilizing early spatial feature maps, and VIB compression for controlling the information content of learned representations. This combination provides a principled basis for improving the reliability of CNN classifiers under both clean and corrupted input conditions.

The practical importance of the topic is also supported by the moderate computational cost of the proposed approach. Since real-world systems often operate under resource constraints, robustness improvements are most valuable when they do not require a substantial increase in model size or inference latency. The PDE-CNN-VIB architecture developed in this dissertation is therefore relevant not only as a theoretical and experimental contribution, but also as a step toward more reliable visual recognition systems for practical applications.

Main Aim of the Work and Considered Problems

The main aim of this dissertation is to improve CNN-based image classification and to develop a robust image classification architecture.

To achieve this aim, the following problems were considered.

1. Analyze and compare existing strategies for applying Information-Theoretic tools in ML, identify their role, effectiveness, and limitations in solving problems related to feature selection, representation learning, model regularization, generalization, calibration, and robustness.
2. Integrate the two-dimensional predefined convolutional layers based on parabolic and hyperbolic PDEs as front-end components in CNN-based image classification, to learn low-level spatial abstractions before conventional feature extraction.

[2] Alemi A. A., Fischer I., Dillon J. V. and Murphy K., "Deep variational information bottleneck", arXiv:1612.00410, 2019.

[3] Aubert, G. and Kornprobst, P., Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations, New York: Springer Verlag, 2002.

3. Combine the Information-theoretic control of the learned representation with the spatial structural regularization extracted by PDE based layers and develop a new hybrid architecture for image classification.
4. Experimentally evaluate the effectiveness of new architecture's generalization, robustness, callibration and computational overhead on several datasets and CNN backbones.

Research objects

The research objects of this dissertation are CNN architectures for image classification, the representation learning processes within these architectures, PDE-based convolutional prior layers and diffusion front-end modules for spatial structural regularization and information-theoretic bottleneck mechanisms for controlling learned representations. These objects are studied in the context of clean and corrupted image classification with attention to generalization, calibration, robustness, and dependence on dataset setting and backbone architecture.

Research methods

The results of the dissertation are based on a combination of mathematical, information-theoretic, numerical, and computational-experimental methods. The theoretical part of the work uses the basic tools of Information Theory, including entropy, conditional entropy, joint entropy, Mutual Information, normalized Mutual Information, cross-entropy, Kullback-Leibler divergence, the IB principle, and the VIB framework. These methods are used to analyze uncertainty, dependence between variables, representation compactness, probabilistic reliability, and the compression-prediction trade-off in ML models.

To study structural regularization in image recognition, methods of PDE and finite-difference approximation are used. In particular, parabolic and hyperbolic PDE-based formulations are considered as mathematical foundations for predefined convolutional layers. For the final PDE-CNN-VIB architecture, the PDE-based prior idea is adapted into a diffusion-based trainable front-end using residual Laplacian updates with learnable nonnegative channel-wise diffusion coefficients. This makes it possible to combine the interpretability of PDE-based smoothing with the flexibility of trainable CNN-based feature extraction and Intormation-theoretic compression mechanism.

The computational part of the dissertation is based on Deep Learning methods for image classification. CNN architectures and their representative backbone families, including ResNet, VGG, and DenseNet, are used to study the behavior of the proposed models under clean and corrupted image conditions.

Experiments

The experimental results are obtained using standard benchmark datasets for image classification and robustness evaluation, including CIFAR-10, CIFAR-100, CIFAR-10-C, and CIFAR-100-C. Clean-data performance is evaluated using classification accuracy, train-test generalization gap, Negative Log-Likelihood, and Expected Calibration Error. Robustness under corrupted inputs is evaluated using corruption-wise accuracy, corruption-category averages, and mean corruption accuracy.

In addition, computational efficiency is studied through the number of trainable parameters, average training time per epoch, inference latency, and robustness-cost trade-off.

Scientific novelty

The scientific novelty of the dissertation consists in the development of PDE-based convolutional prior layers for learning low-level spatial abstractions in CNN-based image classification and their transformation into a lightweight diffusion-based trainable front-end. The proposed front-end uses repeated residual Laplacian diffusion updates with learnable nonnegative channel-wise diffusion coefficients, which allows spatial structural regularization to be integrated into an end-to-end trainable neural architecture. This construction preserves the interpretability of PDE-based smoothing while making it compatible with convolutional feature adaptation, VIB compression, channel reconstruction, and residual blending.

The new developed hybrid PDE-CNN-VIB architecture combines the adapted diffusion-based PDE front-end with a VIB module. The architecture also includes channel reconstruction and learnable residual feature blending, which allows deterministic PDE-enhanced features and stochastic VIB-compressed features to be combined before classification. This forms a new modular approach for robust CNN-based image classification, applicable to different CNN backbones rather than to one specific network only.

Practical significance

The practical significance of results consists in the development of a software and experimental framework for studying robust image classification under clean and corrupted input conditions. The implemented system allows training and evaluating baseline CNN, PDE-CNN, and PDE-CNN-VIB models on standard image classification datasets, including CIFAR-10 and CIFAR-100, as well as their corrupted variants CIFAR-10-C and CIFAR-100-C. The framework makes it possible to compare the effect of diffusion-based PDE regularization and VIB compression under identical training and evaluation conditions.

The software implementation was developed in Python using modern deep learning libraries, mainly PyTorch and Torchvision. It supports the construction of different CNN backbone variants, including ResNet, VGG, and DenseNet, and enables the proposed PDE-CNN-VIB front-end to be combined with these architectures. This modular structure makes the implementation suitable not only for the experiments presented in the dissertation, but also for further research involving other CNN backbones, datasets, or robustness settings.

The developed experimental framework allows the calculation of several evaluation metrics that are important for reliable image classification. In addition to standard classification accuracy, the system computes Negative Log-Likelihood, Expected Calibration Error, train-test generalization gap, corruption-wise accuracy, mean corruption accuracy, and computational-cost indicators such as training time and inference latency. The obtained results can be exported in tabular formats such as CSV, which makes them suitable for further statistical analysis, visualization, and comparison.

The practical value of the new PDE-CNN-VIB architecture is that it can serve as a basis for building more reliable image classification systems in environments where visual inputs may be degraded. Such conditions may occur in autonomous driving, medical image analysis, remote sensing, surveillance, traffic monitoring, industrial visual inspection, and edge or mobile vision systems. In these scenarios, images may be affected by noise, blur, compression artifacts, illumination changes, weather effects, or sensor-related distortions. The proposed architecture is designed to improve the stability of learned representations under such non-ideal conditions while keeping computational overhead moderate.

The results of the dissertation may also be useful for future studies on robust neural-network design. The developed framework demonstrates how structural regularization and information-theoretic compression can be combined in a single trainable architecture. Therefore, the developed approach can be extended to other image classification tasks, larger datasets, alternative PDE formulations, and different information-theoretic regularization mechanisms.

Approbation and Testing of the Obtained Results

The main scientific and experimental results of the dissertation were presented at the International Conference on Computer Science and Information Technologies, CSIT 2025, held in Yerevan, Armenia. The presented work concerned the integration of PDE-based preprocessing with the Variational Information Bottleneck framework for improving CNN-based image classification.

The results of the dissertation were also discussed at the general seminar of the Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia.

Publications

The main scientific and experimental results of the dissertation were published in 5 scientific papers, the list of which is presented at the end of the synopsis.

Scope and Structure of the Work

The dissertation is 125 pages long and consists of an introduction, 4 chapters, a conclusion, and a bibliography comprising 86 references.

Content of the work

The Introduction section substantiates the relevance of the dissertation topic, defines the main objective of the research, outlines the problems under consideration, and presents the scientific novelty, practical significance, and key propositions submitted for defense.

The first chapter has a foundational character. There was introduced the image classification problem, the role of Convolutional Neural Networks, benchmark datasets for image classification, and the Information-Theoretic tools used in the dissertation. Image classification is considered as the task of learning a mapping from an input image to a finite set of class labels. CNNs are discussed as the main architecture for this task, since they preserve the spatial structure of images and learn hierarchical feature representations through local convolutional filters, nonlinear activation functions, pooling operations, and deeper classification layers.

In the chapter the role of benchmark datasets is also described. CIFAR-10 and CIFAR-100 are considered as standard clean-image classification datasets, while CIFAR-10-C and CIFAR-100-C are discussed as corrupted-image benchmarks used to evaluate robustness under distribution shift. This distinction is important because clean test accuracy alone is not sufficient to characterize the reliability of CNN-based classifiers under real-world image degradations such as noise, blur, compression artifacts, and contrast changes.

Then there are presented the main elements of Information Theory used in the dissertation. For a discrete random variable X, entropy is defined as

$$H(X) = -\sum_{x \in X} p(x) \, log \, p(x),$$

where $x \in X$ is a possible outcome of the variable, $p(x)$ is the possibility of observing outcome x. $H(X)$ is interpreted as a measure of uncertainty. Mutual Information between random variables X and Y is given by

$$I(X;Y) = H(X) + H(Y) - H(X,Y),$$

where $H(X)$ and $H(Y)$ are the entropies of the random variables X and Y, $H(X,Y)$ is their joint entropy. $I(X;Y)$ measures the amount of information shared between them. Kullback–Leibler divergence is introduced as a measure of discrepancy between two probability distributions P and Q:

$$D_{\mathrm{KL}}(P \| Q) = \sum_{x} p(x) \log \left(\frac{p(x)}{q(x)} \right),$$

where $p(x)$ and $q(x)$ are the probabilities assigned to the outcome x by P and Q.

Special attention is given to the IB principle, where representation learning is formulated as a trade-off between compression and predictive relevance:

$$\mathcal{L}_{IB} = I(X;T) - \beta I(T;Y),$$

where X represents the input variable or original data, T denotes the compressed representation or bottleneck variable, Y represents the target variable or class label and β is a positive trade-off parameter that controls the balance between compressing the input information and preserving information relevant to the target.

Its variational form, the VIB, introduces a stochastic latent representation z

$$z = \mu + \sigma \odot \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I),$$

which is later used in the new PDE-CNN-VIB architecture. Here μ represents the mean vector of the latent distribution, σ represents the standard deviation vector of the latent distribution, ϵ is a random noise vector sampled from a standard normal distribution with mean 0 and identity covariance matrix I and the $\odot$ denotes element-wise multiplication

Thus, in the first chapter the basic concepts required for the dissertation are established: CNNs as the computational framework, CIFAR and CIFAR-C datasets as the evaluation environment, and Information-Theoretic tools as the mathematical basis for studying uncertainty, representation compactness, calibration, and robustness.

The second chapter is devoted to the analysis and comparison of existing Information-Theoretic strategies used in ML. In particular, the applications of entropy, Mutual Information, Kullback–Leibler divergence, cross-entropy, and bottleneck-based methods in ML were analyzed.

The discussions how Information-Theoretic tools are used in feature selection, dimensionality reduction, decision trees, clustering, metric learning, neural-network regularization, and deep representation learning. Special attention is given to the IB framework and its variational extension. The IB principle describes representation learning as a trade-off between compression and predictive relevance, while the VIB makes this principle applicable to neural networks through stochastic latent representations and KL-divergence regularization.

The limitations of Information-Theoretic methods in modern ML are also analyzed in the chapter. In high-dimensional settings, the estimation of entropy and Mutual Information becomes computationally difficult, and approximation errors may reduce the reliability of information-based objectives. In addition, modern CNN classifiers remain sensitive to distribution shifts, corrupted inputs, and shortcut learning, even when they achieve high accuracy on clean benchmark data.

As a result of the analysis carried out in [1], it is concluded that Information-Theoretic methods, especially the IB and VIB, are powerful tools for improving representation compactness, generalization, and calibration. However, these methods do not explicitly impose spatial structure on image features. This conclusion motivates the use of VIB as the information-compression component of the later PDE-CNN-VIB architecture, where it is combined with a separately developed PDE-based structural mechanism.

The third chapter is devoted to the work, where predefined convolutional layers derived from PDEs were studied for image recognition. The chapter investigates how parabolic and hyperbolic PDE-based operators can be inserted before standard CNN backbones in order to improve low-level feature extraction. Here the structural-prior direction that is later adapted and combined with the VIB in the PDE-CNN-VIB architecture is introduced.

In **Section 3.1**, the motivation for using PDE-based priors in image recognition is discussed. CNNs usually learn their first convolutional filters entirely from data. Although this approach is flexible, it also means that the network must learn low-level structures such as edges, local textures, and smooth regions without any explicit mathematical prior. PDEs, on the other hand, have long been used in image processing for smoothing, diffusion, propagation, and structural regularization. Therefore, finite-difference approximations of PDEs can be interpreted as predefined convolutional operators that impose meaningful local structure before the main trainable CNN begins its feature extraction process.

In **Section 3.2,** the mathematical construction of the predefined PDE layers are presented. Two types of equations are considered. The parabolic case is based on the two-dimensional heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2},$$

which models diffusion-type smoothing. After finite-difference discretization, the update can be written in a convolutional form, where the next feature state is obtained by applying predefined local kernels corresponding to second-order spatial differences.

The hyperbolic case is based on the two-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

Unlike the parabolic update, the hyperbolic formulation includes a second-order temporal dependence and therefore uses information from both the current and previous feature states. This gives the hyperbolic layer a propagation-like behavior, while the parabolic layer has a diffusion-like behavior. Thus, the two PDE types introduce different structural transformations into the early stages of the neural network.

In **Section 3.3**, the general PDE-CNN architecture is described. The input image is first passed through a small number of predefined parabolic or hyperbolic convolutional layers. The resulting feature maps are concatenated and then forwarded to a standard deep neural network for classification. The general architecture is shown in Fig. 3.1. The main advantage of this design is that the PDE-based block acts as a lightweight structural front-end and can be attached to different CNN backbones without redesigning the whole architecture.

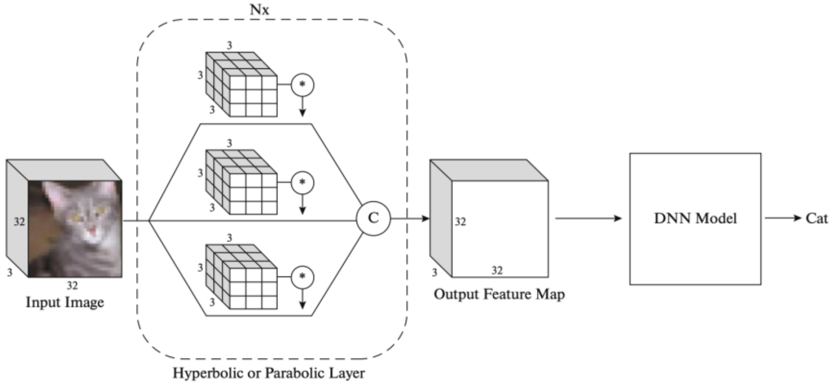

Figure 3.1. General architecture of the PDE-CNN model. The input image is first processed by predefined parabolic or hyperbolic convolutional layers, and the resulting feature maps are passed to the downstream CNN backbone for classification.

In **Section 3.4** the experimental evaluation results on the CIFAR-10 dataset are summarized. The predefined PDE layers were combined with several representative CNN architectures, including ResNet, ResNeXt, VGG, and DenseNet. The experiments showed that the PDE-based layers improve image recognition accuracy in many cases while adding only a very small number of additional trainable parameters. This indicates that the improvement is not caused by a significant increase in model size, but by the structural prior introduced at the beginning of the network.
A compact summary of representative results is given in Table 3.1.

Backbone	Baseline Acc. (%)	Best PDE-Enhanced Acc. (%)
ResNet-18	94.47	95.70
ResNet-34	95.20	95.85
ResNet-50	94.72	95.65
ResNext29	94.90	95.93
VGG-11	92.19	92.65
VGG-13	93.57	94.34
VGG-16	93.85	93.93
DenseNet-121	95.50	96.00

Table 3.1. Representative CIFAR-10 accuracy results for baseline and PDE-enhanced CNNs.

In the chapter there is also a comparison of the PDE-based method with PCFNet, a related predefined-filter approach. The comparison shows that the PDE-based front-end achieves competitive or better accuracy on selected backbones. In addition, the chapter studies the influence of the number of predefined layers. The results indicate that using a large number of PDE layers is not necessary; a small number of parabolic or hyperbolic layers is usually sufficient to obtain the main improvement.

In the final part of the chapter the visualization of the learned predefined kernels is discussed. Hyperbolic kernels often show more directional or polarized behavior, while parabolic kernels tend to resemble Laplacian or cross-shaped smoothing patterns. This confirms that the two PDE families produce different types of low-level structural transformations.

Thus, the results of the third chapter show that PDE-based predefined convolutional layers can serve as lightweight and interpretable structural priors for image recognition. In the dissertation, these results serve as the structural foundation for the next chapter. In Chapter 4, the PDE-based idea is adapted into a diffusion-based trainable front-end and combined with VIB compression in the proposed PDE-CNN-VIB architecture.

In the fourth chapter the main methodological and experimental results of the dissertation are presented, the PDE-CNN-VIB architecture is developed and investigated here. The model combines a diffusion-based PDE front-end, convolutional feature adaptation, a VIB module, residual feature blending, and a CNN classification backbone. The initial idea of combining PDE preprocessing with VIB was introduced in [3], the clean-data evaluation was presented in [4], and the robustness analysis under common corruptions was extended in [5].

First in the chapter, the transition from PDE-based structural priors to the PDE-CNN-VIB architecture are explained. The PDE-based convolutional prior framework used predefined parabolic and hyperbolic PDE layers as front-end components for CNN image recognition. This idea is modified: instead of directly using the full predefined parabolic/hyperbolic construction, the model uses a lightweight diffusion-based parabolic PDE front-end. This front-end applies repeated residual Laplacian updates with learnable nonnegative channel-wise diffusion coefficients. Thus, the PDE component preserves the interpretation of heat-equation-type smoothing while becoming better suited for integration with stochastic representation compression.

The main PDE update used in the architecture has the form

$$u_c^{(r+1)} = u_c^{(r)} + \lambda_c \Delta u_c^{(r)},$$

where $u_c^{(r)}$ is the feature map of channel $c$at diffusion step r, Δ is the discrete Laplacian operator, and λ_c is a learnable nonnegative diffusion coefficient. This update corresponds to a parabolic diffusion process and is used to stabilize early feature maps and reduce sensitivity to high-frequency perturbations.

The general architecture of the proposed PDE-CNN-VIB model is shown in Fig. 4.1.

After the PDE front-end and convolutional feature adaptation stage, the VIB module is applied to the intermediate feature maps. The VIB module produces the parameters of a Gaussian latent representation using two 1×1convolutional heads:

$$\mu = W_\mu * g + b_\mu, \quad \log \sigma^2 = W_\sigma * g + b_\sigma.$$

The stochastic representation is sampled by the reparameterization rule

$$t = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I).$$

The sampled tensor is reconstructed to the original channel dimension and combined with the prebottleneck feature map f using a learnable residual blend, using α learnable parameter to control the strength of the feature map f:

$$h = \alpha t + (1 - \alpha) f.$$

The resulting representation is passed to a CNN backbone, such as ResNet-18, VGG-16, or DenseNet-121, and then to the final classifier. The training objective combines the standard cross-entropy classification loss with a KL-divergence regularization term:

$$\mathcal{L} = \mathcal{L}_{CE}(y, logits) + \beta\, D_{KL}\big(q(\,t \mid x\,) \parallel p(t)\big).$$

Here, β controls the strength of information compression. The role of the VIB module is to suppress redundant and task-irrelevant information while preserving the information necessary for classification.

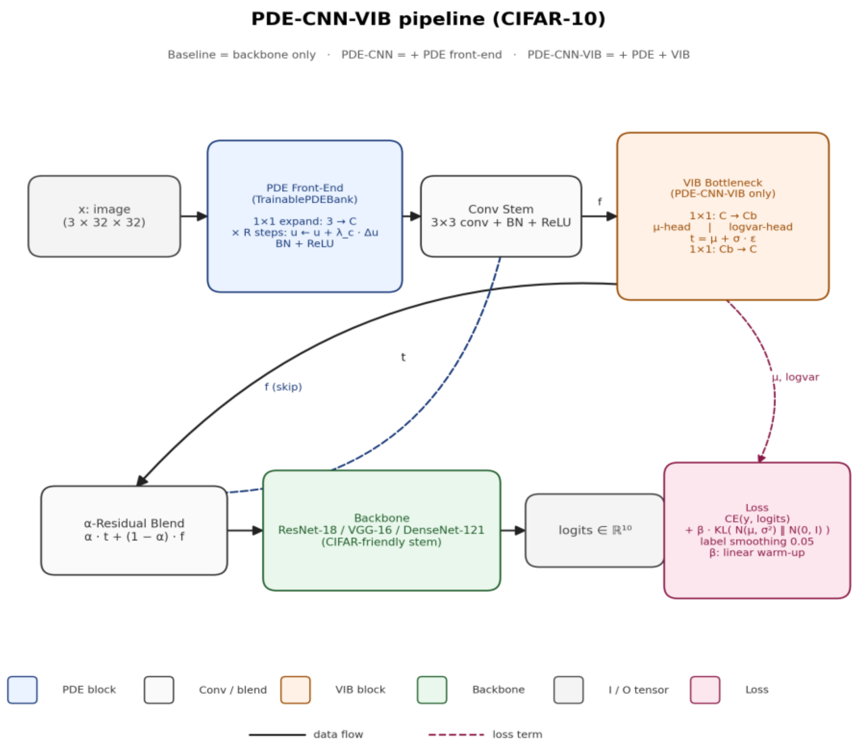

Fig. 4.1. PDE-CNN-VIB architecture used in CIFAR-based experiments. The input image is processed by a diffusion-based PDE front-end and a convolutional stem. The resulting feature representation is passed through a VIB branch and also preserved through a skip connection. The reconstructed VIB representation and the skip representation are combined by a learnable residual blend before being forwarded to the CNN backbone.

The clean-data experiments first evaluate the proposed model on CIFAR-10. In the initial experiment, three variants are compared: a baseline CNN, a PDE-enhanced model, and the full PDE-VIB model. The results show that the PDE-VIB model gives the best balance between classification accuracy and probabilistic reliability. In particular, it improves clean accuracy and reduces Expected Calibration Error compared with the baseline and PDE-only variants. This indicates that PDE preprocessing improves feature extraction, while VIB compression improves representation quality and calibration.

To study longer training behavior, an additional 100-epoch experiment was conducted using ResNet-18 as the base architecture. The results are shown in Table 4.1.

Model	Train Acc. (%)	Test Acc. (%)	Gap (%)	NLL	ECE	CIFAR-10-C mCA (%)
ResNet-18 Baseline	99.40	91.77	7.63	0.3281	0.0369	69.77
ResNet-18 PDE-CNN	99.63	92.59	7.04	0.2998	0.0385	75.27
ResNet-18 PDE-CNN-VIB	99.22	**94.66**	**4.56**	**0.2113**	**0.0235**	**78.26**

Table 4.1. Final clean CIFAR-10 and CIFAR-10-C results for the 100-epoch ResNet-18 experiment.

The results show that the PDE-CNN-VIB model achieves the highest test accuracy, the smallest train-test gap, the lowest NLL, and the lowest ECE. Although its training accuracy is slightly lower than that of PDE-CNN, its test accuracy is substantially higher. This indicates that the VIB component does not simply increase memorization, but improves generalization by regulating the information flow through the network.

Then robustness under common image corruptions is evaluated. Robustness is measured using mean corruption accuracy:

$$\mathrm{mCA} = \frac{1}{15}\sum_{i=1}^{15} \mathrm{Acc_i}\,,$$

where Acc_i is the classification accuracy for the i-th corruption type. The CIFAR-10-C and CIFAR-100-C benchmarks are used at severity level 3.

Examples of clean and corrupted images used in the robustness evaluation are shown in Fig. 4.2. The figure demonstrates how the same clean CIFAR image can be transformed by different corruption types, including noise, blur, frost, and impulse noise. Although the semantic content of the image remains the same, the visual appearance is degraded: local structures may become distorted, textures may be weakened, color distributions may change, and important object boundaries may become less clear.

Therefore, corrupted-image evaluation is more challenging than standard clean-data testing. In this setting, the model is expected not only to classify clean images correctly, but also to preserve stable prediction behavior when the input is affected by common visual corruptions. This makes corrupted-image benchmarks important for evaluating the robustness and practical reliability of image classification models.

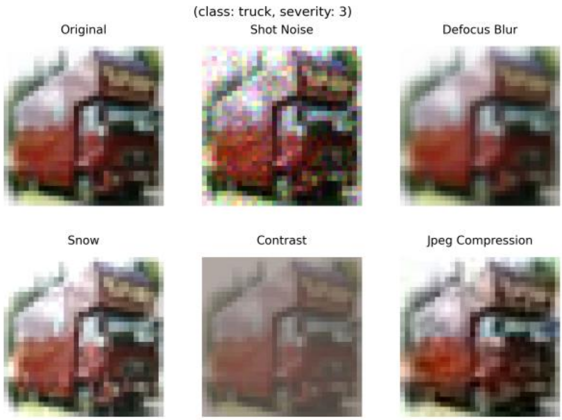

Fig. 4.2. Examples of clean and corrupted CIFAR images used in the robustness evaluation.

First, the architecture is tested under a 20-epoch training protocol with three representative CNN backbones: ResNet-18, VGG-16, and DenseNet-121. The results are summarized in Table 4.2.

Backbone	Variant	Accuracy (%)	NLL	ECE	mCA (%)
ResNet-18	Baseline	86.23	0.4896	0.0323	60.80
ResNet-18	PDE-CNN	88.00	0.4042	0.0229	66.09
ResNet-18	PDE-CNN-VIB	**89.44**	**0.3432**	**0.0206**	**71.24**
VGG-16	Baseline	84.13	0.5146	0.0203	62.16
VGG-16	PDE-CNN	84.52	0.4998	0.0187	58.86
VGG-16	PDE-CNN-VIB	**87.15**	**0.4121**	**0.0118**	**67.24**
DenseNet-121	Baseline	85.65	0.3632	0.0333	65.48
DenseNet-121	PDE-CNN	89.56	0.3415	0.0283	68.84
DenseNet-121	PDE-CNN-VIB	**91.07**	**0.3284**	**0.0237**	**69.28**

Table 4.2. Initial 15-epoch CIFAR-10 and CIFAR-10-C results across different CNN backbones.

The table shows that PDE-CNN-VIB gives the best overall result within each backbone family. This is important because it shows that the proposed approach is not restricted to a single CNN architecture. The VGG-16 case is especially informative: PDE-CNN alone reduces mCA, while PDE-CNN-VIB improves it substantially. This confirms that the VIB module plays an important stabilizing role and helps make the PDE-enhanced representation more robust.

The 100-epoch ResNet-18 experiment also confirms the robustness advantage of PDE-CNN-VIB. In this setting, the model improves CIFAR-10-C mCA from 69.77% for the baseline to 78.26%. The strongest gains are observed for corruptions that affect local image structure, especially Gaussian noise, impulse noise, glass blur, zoom blur, snow, and frost. Category-wise analysis shows that PDE-CNN-VIB improves all major corruption families: noise, blur, weather, and digital distortions.

To verify that the method is not limited to CIFAR-10, the model is also evaluated on CIFAR-100 and CIFAR-100-C. CIFAR-100 is more difficult because it contains 100 classes instead of 10, while preserving the same 32×32 image format. The results are shown in Table 4.3.

Model	Clean Acc. (%)	NLL	ECE	mCA (%)
ResNet-18 Baseline	67.09	1.2892	0.0519	41.18
ResNet-18 PDE-CNN	66.91	1.2711	0.0482	43.86
ResNet-18 PDE-CNN-VIB	**73.05**	**0.8023**	**0.0321**	**45.66**

Table 4.3. CIFAR-100 and CIFAR-100-C results for ResNet-18-based models.

The CIFAR-100 results show that PDE-CNN-VIB improves clean accuracy from 67.09% to 73.05%, while also reducing NLL and ECE. The CIFAR-100-C mCA also increases from 41.18% to 45.66%. Thus, the proposed architecture improves both clean-data behavior and corrupted-data robustness even when the classification problem becomes more fine-grained and difficult.

The computational cost of the new architecture is also considered in the chapter. The PDE-CNN-VIB model introduces only a small increase in the number of parameters. In the ResNet-18-based setting, the parameter count increases from 11.17 million to 11.20 million. The inference latency increases only slightly, from 1.463 ms per image to 1.516 ms per image. In the 100-epoch experiment, the average epoch time increases from 111.17 s for the baseline to 120.84 s for PDE-CNN-VIB. Therefore, the robustness and calibration improvements are achieved with moderate computational overhead.

In the final part of the chapter possible practical use-case scenarios are discussed. Since the model improves robustness under noise, blur, compression artifacts, and other non-ideal image conditions, it may be useful in autonomous driving, medical image analysis, remote sensing, surveillance, industrial visual inspection, and edge/mobile vision systems. In all these domains, visual inputs may be degraded, and therefore classification reliability depends not only on clean accuracy, but also on robustness and calibration.

Thus, the fourth chapter presents the main contribution of the dissertation. The previously studied PDE-based structural prior is adapted into a lightweight diffusion-based trainable front-end, and this front-end is integrated with VIB-based stochastic compression in the PDE-CNN-VIB architecture. The obtained results show that the proposed architecture improves clean classification behavior, reduces calibration error, increases mean corruption accuracy on CIFAR-10-C and CIFAR-100-C, and remains applicable across several CNN backbones.

Main results of the work

1. A comprehensive analysis of the current implications of Information-Theoretic tools and methods in ML was conducted, which highlighted the critical role of Information-Theoretic methods in controlling representation quality, reducing redundancy, and improving generalization, but also revealed that such methods alone do not explicitly impose spatial structure on visual representations [1].
2. A PDE-based convolutional prior was introduced for CNN image classification by adding two-dimensional predefined convolutional layers based on parabolic and hyperbolic PDEs as front-end layers before standard CNN backbones. The method showed accuracy

improvements across several architectures, while preserving low computational overhead in terms of trainable parameters and FLOPs. [2].

3. A new hybrid PDE-CNN-VIB architecture was developed by integrating the adapted diffusion-based PDE front-end with convolutional feature adaptation and a VIB module. The resulting architecture combines two complementary mechanisms: PDE-based structural regularization for stabilizing early spatial feature maps and VIB-based information-theoretic compression for suppressing redundant and task-irrelevant information. This integration forms a new CNN-based model for robust image classification under clean and corrupted input conditions [3].
4. The effectiveness of the new PDE-CNN-VIB architecture was substantiated through experiments on multiple benchmark datasets and CNN backbones. Results provide empirical evidence that the new model is not dependent on a particular benchmark or underlying architecture. The experiments showed improved clean-data behavior, calibration, generalization, and corruption robustness, while keeping the inference overhead nearly unchanged compared with the baseline models [4, 5].

List of the author's publications

1. M. Haroutunian, G. Gharagyozyan, Information Theory Tools and Techniques to Overcome Machine Learning Challenges, Mathematical Problems of Computer Science 63, 25-41, 2025. doi: 10.51408/1963-0129
2. V. Sahakyan, V. Melkonyan, G. Gharagyozyan, A. Avetisyan, Enhancing Image Recognition with Pre-Defined Convolutional Layers Based on PDEs, ISSN 0361-7688, Programming and Computer Software, 2024, Vol. 49, Suppl. 3, pp. 192-197. © Pleiades Publishing, Ltd., 2024. doi: 10.1134/S0361768823030088
3. G. Gharagyozyan, Improving CNN Generalization with PDE Preprocessing and the Variational Information Bottleneck, In the proceedings of CSIT Conference 2025, 145-147, Yerevan, Armenia, 2025. doi: 10.51408/csit2025_35
4. G. Gharagyozyan, A PDE-Based Convolutional Neural Network with Variational Information Bottleneck: Experimental Evaluation and Generalization Analysis, Mathematical Problems of Computer Science 64, 37-46, 2025. doi: 10.51408/1963-0138
5. G. Gharagyozyan, Improving Robust Image Classification Under Common Corruptions: A PDE-Regularized Variational Information Bottleneck Network, ISSN 3067-2384, World Journal of Sensors Network Research, 2026, Vol. 3, pp. 1-10. https://doi.org/10.63620/MKWJSNR.2026.1051.

ВНЕДРЕНИЕ ИНСТРУМЕНТОВ ТЕОРИИ ИНФОРМАЦИИ В АРХИТЕКТУРУ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

Абстракт

В условиях быстрого развития цифровых технологий, искусственного интеллекта и машинного обучения задачи анализа и классификации изображений приобретают особую научную и прикладную значимость. Цифровые изображения широко используются в медицинской диагностике, автономных системах, дистанционном зондировании, видеонаблюдении, промышленном контроле качества и других областях. В таких приложениях модели должны быть надежными не только на чистых тестовых данных, но и в ситуациях, когда изображения содержат шум, размытие, артефакты сжатия, изменения освещения или другие искажения.

Сверточные нейронные сети (CNN) являются одним из основных инструментов классификации изображений, поскольку способны извлекать иерархические признаки из визуальных данных. Однако высокая точность на стандартных эталоных наборах не всегда гарантирует устойчивость модели в реальных условиях. При изменении распределения входных данных CNN-модели могут терять точность, демонстрировать слабую калибровку и выдавать чрезмерно уверенные ошибочные предсказания. Поэтому разработка более устойчивых, хорошо калиброванных и обобщающих CNN-классификаторов остается актуальной задачей.

Актуальность данной диссертации связана с исследованием двух взаимодополняющих направлений. Первое направление состоит в использовании сверточных и диффузионных модулей, основанных на дифференциальных уравнениях в частных производных (PDE), для пространственной структурной регуляризации ранних признаков. Такие компоненты позволяют учитывать локальную структуру изображения и стабилизировать низкоуровневые представления. Второе направление связано с применением информационно-теоретических методов, в частности Variational Information Bottleneck (VIB), для управления количеством информации, передаваемой через модель. Это позволяет уменьшить избыточные и нестабильные признаки, улучшить обобщающую способность, калибровку и устойчивость классификатора.

Таким образом, диссертация направлена на разработку и исследование гибридной архитектуры, объединяющей PDE-регуляризацию и информационно-теоретическое сжатие представлений для более надежной классификации изображений.

Основные результаты работы

1. Проведен многосторонний анализ современных применений инструментов и методов теории информации в машинном обучении, который подчеркнул решающую роль методов теории информации в контроле качества представлений, снижении избыточности и улучшении обобщающей способности, но также показал, что сами по себе такие методы не навязывают визуальным представлениям явной пространственной структуры [1].

2. Для классификации изображений с помощью CNN был предложен конволюционный априорный распределитель на основе дифференциальных уравнений в частных производных (PDE): перед стандартными базовыми сетками CNN были добавлены двумерные заранее заданные конволюционные слои, основанные на параболических и гиперболических PDE. Данный метод продемонстрировал повышение точности на нескольких архитектурах, сохранив при этом низкие вычислительные затраты с точки зрения количества обучаемых параметров и количества операций с плавающей запятой (FLOP). [2].

3. Была разработана новая гибридная архитектура PDE-CNN-VIB путем интеграции адаптированного фронт-энда на основе диффузионного уравнения с конволюционной адаптацией признаков и модулем VIB. Получившаяся архитектура сочетает в себе два взаимодополняющих механизма: структурную регуляризацию на основе уравнения PDE для стабилизации ранних пространственных карт признаков и информационно-теоретическое сжатие на основе VIB для подавления избыточной и нерелевантной для задачи информации. Такая интеграция формирует новую модель на основе CNN для надежной классификации изображений как при чистом, так и при поврежденном входном сигнале. [3].

4. Эффективность новой архитектуры PDE-CNN-VIB была подтверждена в ходе экспериментов с использованием нескольких тестовых наборов данных и базовых CNN-архитектур. Результаты экспериментов свидетельствуют о том, что новая модель не зависит от конкретного тестового набора или базовой архитектуры. Эксперименты продемонстрировали улучшение характеристик при работе с чистыми данными, калибровку, способность к обобщению и устойчивость к искажениям данных, при этом накладные расходы на вычисления остались практически неизменными по сравнению с базовыми моделями [4, 5].

ԻՆՖՈՐՄԱՑԻԱՅԻ ՏԵՍՈՒԹՅԱՆ ԳՈՐԾԻՔԱԿԱԶՄԻ ՆԵՐԴՐՈՒՄԸ ԽՈՐՔԱՅԻՆ ՆԵՅՐՈՆԱՅԻՆ ՑԱՆՑԻ ՃԱՐՏԱՐԱՊԵՏՈՒԹՅՈՒՆՈՒՄ՝ ՊԱՏԿԵՐՆԵՐԻ ԴԱՍԱԿԱՐԳՄԱՆ ԱՐԴՅՈՒՆԱՎԵՏՈՒԹՅԱՆ ԲԱՐՁՐԱՑՄԱՆ ՆՊԱՏԱԿՈՎ

Ամփոփում

Թվային տեխնոլոգիաների, արհեստական բանականության և մեքենայական ուսուցման արագ զարգացման հետ մեկտեղ պատկերների վերլուծության և դասակարգման խնդիրները ձեռք են բերում առանձնահատուկ գիտական և գործնական նշանակություն: Թվային պատկերները լայնորեն կիրառվում են բժշկական ախտորոշման, ավտոնոմ համակարգերի, հեռազննման, տեսահսկման, արդյունաբերական որակի վերահսկման և այլ ոլորտներում: Այսպիսի կիրառություններում մոդելները պետք է լինեն հուսալի ոչ միայն մաքուր փորձարկման տվյալների դեպքում, այլև այնպիսի իրավիճակներում, երբ պատկերները պարունակում են աղմուկ, սեղմման աղավաղումներ, լուսավորության փոփոխություններ կամ այլ աղավաղումներ:

Կոնվոլյուցիոն նեյրոնային ցանցերը (CNN) պատկերի դասակարգման հիմնական գործիքներից են, քանի որ կարող են տեսողական տվյալներից հանել հիերարխիկ առանձնահատկություններ: Սակայն ստանդարտ փորձարարական տվյալների բազմությունների վրա բարձր ճշգրտությունը միշտ չէ, որ երաշխավորում է մոդելի հուսալի աշխատանքը իրական աշխարհի պայմաններում: Երբ մուտքային տվյալների բաշխումը փոխվում է, CNN մոդելները կարող են կորցնել ճշգրտությունը, ցուցաբերել թույլ կալիբրացում և տալ չափազանց ինքնավստահ սխալ կանխատեսումներ: Ուստի ավելի դիմացկուն, լավ կալիբրացված և կայուն CNN դասակարգիչների մշակումը շարունակում է մնալ արդիական խնդիր:

Այս աշխատանքի արդիականությունը կայանում է երկու համալրող ուղղությունների ուսումնասիրության մեջ: Առաջին ուղղությունը ներառում է կոնվոլյուցիայի և դիֆուզիայի

մոդուլների կիրառումը, որոնք հիմնված են մասնակի ածանցյալներով դիֆերենցիալ հավասարումների (PDE) վրա՝ վաղ փուլային առանձնահատկությունների տարածական կառուցվածքային կարգավորման համար: Երկրորդ ուղղությունը ներառում է ինֆորմացիայի տեսության մեթոդների, մասնավորապես Վարիացիոն Ինֆորմացիոն Խցանի (Variational Information Bottleneck, VIB), կիրառումը մոդելի միջոցով փոխանցվող տեղեկատվության քանակը վերահսկելու համար: Սա օգնում է նվազեցնել կրկնվող և անկայուն հատկանիշները և բարելավել դասակարգչի ընդհանրացման ունակությունը, կալիբրավորումը և կայունությունը:

Այսպիսով, աշխատանքը նպատակ ունի մշակել և ուսումնասիրել հիբրիդային ճարտարապետություն, որը համատեղում է PDE կանոնավորումը և ինֆորմացիայի տեսության վրա հիմնված ներկայացման սեղմումը՝ ավելի հուսալի պատկերների դասակարգում ապահովելու համար:

Աշխատանքի հիմնական արդյունքները

1. Կատարվել է մեքենայական ուսուցման մեջ ինֆորմացիայի տեսության ժամանակակից գործիքակազմի կիրառության համապարփակ հետազոտություն և վերլուծություն, ընդգծվել է դրանց կարևոր դերը ներկայացումների որակի վերահսկման և ընդհանրացման բարելավման խնդիրներում, միևնույն ժամանակ ցույց է տրվել, որ տվյալ մեթոդները ինքնուրույն չեն պարտադրում տեսողական ներկայացումների վրա հստակ տարածական կառուցվածք [1]:

2. Պատկերների դասակարգման համար ներդրվել է մասնակի ածանցյալներով դիֆերենցիալ հավասարումների վրա հիմնված նախահայտարարված նախաշերտ, որը մի շարք ճարտարապետություններում ցուցաբերել է ճշգրտության բարելավումներ, ցածր պահելով հաշվողական բարդությունը [2]:

3. Մշակվել է նոր հիբրիդային PDE-CNN-VIB ճարտարապետություն, որը համատեղում է ադապտացված դիֆուզիոն PDE նախաշերտը, կոնվոլյուցիոն հատկանիշների հարմարեցումը և Վարիացիոն Ինֆորմացիայի Խցանի մոդուլը: Ճարտարապետությունը համատեղում է վաղաժամ տարածական առանձնահատկությունների PDE կանոնավորումը և ինֆորմացիոն-տեսական սեղմումը՝ ավելորդ և ոչ համապատասխան տեղեկատվությունը ճնշելու նպատակով [3]:

4. Նոր ճարտարապետության արդյունավետությունը հաստատվել է փորձարկումներով՝ օգտագործելով տարբեր տվյալների հավաքածուներ և CNN հիմքեր: Արդյունքները ցույց են տալիս, որ նոր մոդելը կախված չէ կոնկրետ տվյալների հավաքածուից կամ CNN հիմքից: Փորձարկումները ցույց են տվել բարելավում մաքուր տվյալների վրա ճշգրտությունում, կալիբրացիայի գործընթացում, ընդհանրացման ունակությունում և աղավաղումների նկատմամբ կայունությունում, մինչդեռ հաշվարկային բարդությունը մնացել է գրեթե անփոփոխ համեմատած հիմնական մոդելների [4, 5]: