ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

Hyper-Ada: Active Data Augmentation via High-Order Feature Refinement with Adaptive Hypergraphs

Sidan Hou[a], Yu Xiang[b,*], Wei Wang[c], Min Li[d]

School of Information Science and Technology, Yunnan Normal University, Kunming, China
[a]2729229971@qq.com, [b]iamlionx@126.com, [c]1848907205@qq.com, [d]2243020012@qq.com
**Corresponding author*

Abstract: *Data augmentation (DA) is a critical technique for addressing data scarcity in deep learning. However, traditional random augmentation methods are inefficient, while active learning (AL) often overlooks the high-order correlations among samples. To address these limitations, we propose the Progressive Active Data Augmentation (PADA) framework, which applies the intelligent selection principles of AL to the data augmentation process. Within this framework, we designed the core selection strategy, Hyper-Ada. This strategy leverages an Adaptive Hypergraph Convolution (AdaHGConv) network to refine new features embedded with high-order relations, derived from the global context of all augmented samples. We innovatively combine the model's prediction "Certainty" with our proposed "Representation Shift"—the magnitude of feature change before and after refinement—as the selection criterion to identify high-quality "anchor" samples. Experiments on CIFAR-10, SVHN, and CIFAR-100 demonstrate that Hyper-Ada significantly outperforms supervised baselines and traditional AL methods, particularly in data-sparse scenarios. This validates the efficacy of guiding data augmentation through high-order feature refinement.*

Keywords: *Data Augmentation, High - order Features, Active Learning, Adaptive Hypergraph, Image Classification*

1. Introduction

The remarkable performance of deep neural networks in computer vision tasks is highly dependent on the availability of large-scale annotated datasets[1][2]. However, in many real-world applications, the high cost of data annotation constitutes a significant bottleneck, limiting model performance and applicability. Data Augmentation (DA), which expands datasets through artificial transformations, is a common technique for enhancing model generalization. Yet, conventional stochastic augmentation methods are often indiscriminate, producing a large volume of redundant or non-beneficial samples and thus reducing training efficiency.

Concurrently, Active Learning (AL) aims to mitigate annotation costs by intelligently selecting the most informative samples for labeling. However, traditional AL strategies exhibit two primary limitations: First, they primarily focus on selecting samples from an unlabeled pool for costly manual annotation. Second, their selection criteria—such as uncertainty[3] or clustering-based diversity[4]—largely evaluate samples in isolation, ignoring the complex internal structure of the dataset as a whole and the high-order relations among samples.

This inspired us to investigate whether the "intelligent selection" paradigm of AL could be applied to the "cost-free" process of data augmentation, and whether an advanced selection strategy capable of comprehending high-order inter-sample relationships could be designed.

To achieve this, we propose a new framework named "Progressive Active Data Augmentation (PADA)." During the model training process, this framework periodically selects a batch of the most valuable samples from a large candidate pool generated by strong data augmentation, thereby optimizing the training set.

At the core of the PADA framework, we further introduce a novel selection strategy named Hyper-Ada. Inspired by the principles of adaptive hypergraph networks[5] Hyper-Ada moves beyond treating candidate samples as independent points. It processes the initial features of all candidates as a global

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

context, leveraging Adaptive Hypergraph Convolution to dynamically discover latent high-order semantic correlations. Based on these correlations, it refines each sample's feature representation to encode this structural information.

Building on this, we innovatively propose two core metrics to evaluate the value of augmented samples. The first is "Representation Shift," defined as the L2 distance of a sample's feature vector before and after refinement. We hypothesize that a greater shift magnitude signifies a more critical role for the sample within the data manifold and a stronger high-order correlation with other samples. Contrary to traditional AL, which prioritizes "uncertain" samples, our experiments revealed that combining this metric with the model's prediction "confidence" is a more effective strategy. Our ideal samples are those that the model can already recognize with high confidence (High Certainty) but that also play a significant role in the global data structure (High Representation Shift). These samples can be regarded as "high-quality anchors" for each class. Adding them to the training set most effectively helps the model to consolidate and stabilize its decision boundary.

1.1 Main Contributions

We conducted comprehensive experimental validation on three classic image classification benchmarks (CIFAR-10, SVHN, and CIFAR-100), particularly in simulated data-sparse scenarios (e.g., using only 20% of the original training data). The results demonstrate that the Hyper-Ada strategy within the PADA framework significantly outperforms standard supervised baselines and traditional "uncertainty + clustering" active selection methods. The main contributions of this paper are summarized as follows:

We propose PADA, a novel framework that applies active learning principles to optimize the data augmentation process.

We design Hyper-Ada, an advanced active selection strategy based on adaptive hypergraphs and high-order feature refinement.

We demonstrate through extensive experiments that in data-sparse scenarios, understanding and leveraging high-order inter-sample relationships to guide data augmentation can effectively enhance the performance of deep learning models.

2. Related Work

2.1 Data Augmentation

Data augmentation is a cornerstone for mitigating the data dependency of deep neural networks. Early efforts primarily focused on simple geometric transformations and color jittering, such as random flipping, cropping, rotation, and adjustments to brightness and contrast[1]. While simple and effective, these methods possess a limited transformation space. To further enrich data diversity, researchers introduced region-erasure methods like Cutout[6] and Random Erasing[7], which compel the model to focus on more comprehensive contextual information by randomly occluding parts of an image.

Furthermore, sample-mixing strategies, such as Mixup[8] and CutMix[9], effectively expand the training distribution by performing linear interpolation at both the sample and label levels. These have been proven to enhance model generalization and robustness against adversarial attacks. To overcome the limitations of manually designing augmentation strategies, automated data augmentation (AutoAugment)[10] and its variants, like RandAugment[11], employ reinforcement learning or grid search to automatically discover optimal combinations of augmentation policies.

Despite these methods significantly enriching the DA toolkit, their common characteristic is the universality and stochasticity of the policies. That is, once an augmentation strategy is defined, it is applied randomly and indiscriminately to all training samples, without considering the model's specific "demand" for certain types of augmented data at particular training stages. In contrast, our PADA framework introduces a selection mechanism that aims to actively and purposefully select the most valuable samples for the current model from a randomly generated augmentation pool. This transforms data augmentation from an "open-supply" process into an "on-demand optimization" process.

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

2.2 Active Learning

Active learning (AL) aims to minimize annotation costs by querying the most informative unlabeled samples. The core of AL lies in its sample selection strategy. The most classic category of strategies is Uncertainty-based Sampling[3], which selects samples where the model has the lowest prediction probability (Least Confidence), the highest prediction entropy (Max Entropy), or the smallest difference between the top predicted class probabilities (Margin Sampling). The intuition behind these methods is that the model learns the most from samples near its decision boundary.

However, relying solely on uncertainty may lead to a selection of samples lacking diversity. To address this, researchers proposed Diversity-based Sampling. For instance, the Core-set approach[4] attempts to find a core subset that best represents the entire dataset distribution. Another common method involves using clustering algorithms to select samples from different clusters, ensuring broad coverage.

In recent years, research has begun to integrate both uncertainty and diversity, exploring more complex selection criteria such as expected model change[12] or gradient information[13]. Nonetheless, most existing AL methods still suffer from two limitations: 1) They are primarily applied to scenarios requiring expensive manual annotation. 2) Their "diversity" metrics are often confined to pairwise relations between samples, failing to capture more complex, high-order structures formed by groups of samples. Our work overcomes both limitations: first, we apply the AL philosophy to the "cost-free" candidate pool of data augmentation; second, our Hyper-Ada strategy explicitly models the high-order relationships between samples, achieving a deeper understanding of the data's intrinsic structure.

2.3 Graph and Hypergraph Learning

Graph Neural Networks (GNNs)[14][15]have emerged as powerful tools for processing structured data, learning node representations by passing and aggregating messages along the graph structure. However, standard graphs can only represent binary (pairwise) relationships between nodes. In many real-world scenarios, the relationships among samples are far more complex than pairwise connections.

Hypergraphs[16][17] provide a more powerful framework, wherein a single "hyperedge" can connect an arbitrary number of nodes, thereby naturally representing multi-way, high-order relationships. In recent years, hypergraph learning, particularly Hypergraph Neural Networks, has demonstrated significant potential in computer vision[18] and other domains. For example, adaptive hypergraph networks[5] propose a method for dynamically generating hyperedges. This method can discover latent high-order semantic correlations based on the global context of the input data and refine each node's feature representation via hypergraph convolution.

Introducing graph or hypergraph structures into active learning is an emerging research direction[19]. These methods typically leverage the graph structure to measure uncertainty and representativeness simultaneously. However, to the best of our knowledge, our work is the first to utilize the "adaptive" hypergraph concept to refine high-order features of augmented samples and to innovatively use "Representation Shift" and "Certainty" as selection criteria. Our Hyper-Ada strategy is built upon this cutting-edge idea, aiming to achieve unprecedented intelligent data augmentation by modeling the high-order intrinsic structure of the data.

3. Methods

3.1 Progressive Active Data Augmentation (PADA) Framework

The traditional data augmentation process is static: once a set of augmentation strategies is determined, they are randomly applied to the data throughout the training process. We believe that a more effective paradigm should be dynamic and adaptive, that is, the model should actively select the most beneficial reinforcement samples for itself based on its learning status at different training stages. Based on this, we have proposed the PADA framework, whose process is shown in Figure 1. The specific steps are as follows:

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

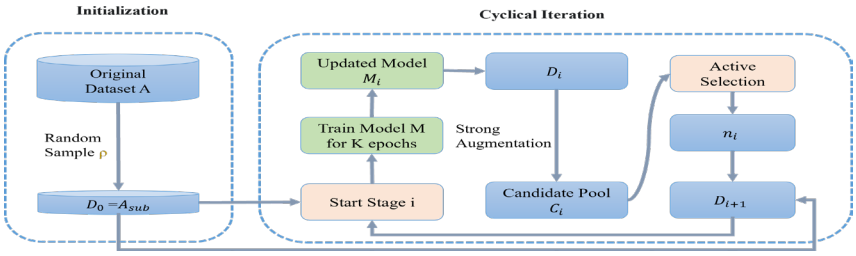

Figure 1: PADA framework

1) Initialization: Given an original labeled training set A, we first randomly sample a subset A_{sub} from it based on the preset ratio ρ (subset-ratio) as our base training set. This step aims to simulate scenarios where data is scarce. The initial training dataset D_0 is set to A_{sub}.

2) Cyclical Iteration: The entire training process is divided into multiple stages, each of which contains K epochs (active-interval). For the i-th stage:

a) Training: Conduct K epochs of standard supervised training on the deep learning model M using the current training dataset D_{i} to obtain the updated model M_{i}.

b) Candidate Pool Generation: Apply a set of predefined, powerful random data augmentation transformations (such as RandAugment [11]) to the training set D_{i} of the current stage to generate a large, temporary augmentation candidate pool C_{i}.

c) Active Selection: Input the updated model M_{i} and candidate pool C_{i} into our Hyper-Ada selection module (see Section 3.2 for details). This module will screen out n enhancement samples that are considered the most valuable from C_{i} and denote them as n_{i}. Here, n is usually a fixed proportion (e.g. 10%) of the size of A_{sub}.

d) Dataset Update: Construct a new dataset $D_{\text{i+1}}$ for the next stage of training. In our implementation, we adopt the substitution strategy, namely $D_{\text{i+1}} = A_{\text{sub}} \cup n_{\text{i}}$. This means that the enhanced samples at each stage are de novo screened based on the latest model status, ensuring the dynamic adaptability of the strategy.

3) Termination: Repeat step 2 until the preset total training epochs are reached. The PADA framework transforms the traditional, one-off data augmentation process into a periodic "perception-decision-making - optimization" closed loop closely coupled with model training, thereby achieving the intelligence of data augmentation.

3.2 Hyper-Ada: Adaptive Hypergraph Filtering Strategy

The Hyper-Ada filtering strategy is the core driving force of the PADA framework, and its architecture is shown in Figure 2. Its objective is to provide a profound and effective decision-making basis for the active screening in step 2.c). The design of this strategy is based on a core assumption: the most valuable enhanced samples are those that play a key role in the overall structural relationship of the data. To capture this complex structure that transcends individual and paired relationships, we introduce an adaptive hypergraph convolutional network.

For each sample x_{j} in the enhanced candidate pool C_{i}, we first extract the information of two fundamental dimensions using the current model M_{i}:

Initial Feature: We input x_{j} into the backbone network of M_{i} and extract its feature vector f_{j} before the global average pooling layer. This vector $\mathbf{f}_j \in \mathbb{R}^D$ is the original, low-order semantic encoding of the content of x_{j} by the model.

Prediction Confidence: We perform a complete forward propagation on x_{j} to obtain its prediction probability distribution p_{j} across C categories. The confidence level c_{j} of this sample is defined as the maximum predicted probability, that is, $c_{\text{j}} = \max(p_{\text{j}})$.

After completing this step, for M samples in the candidate pool, we obtain the initial feature set F = f_1,... f_{M} and the confidence set C = c_1,... c_{M}.

This is the technical core of the Hyper-Ada strategy. Instead of conducting the analysis directly on the initial feature F, we first refined it through an adaptive hypergraph convolutional network

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

(AdaHGConv) [5]to inject high-order, context-dependent structural information.

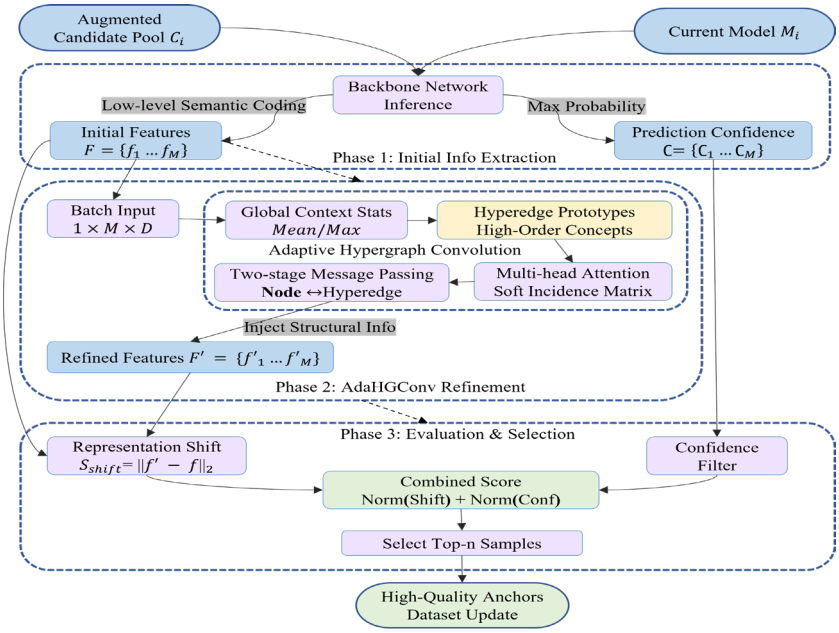

Figure 2: Hyper-Ada framework.

Specifically, we input the entire initial feature set F (an M x D matrix) as a complete batch (1 x M x D) into the AdaHGConv module. This module performs two key operations internally:

1) Adaptive Hyperedge Generation: The module first calculates the global statistics (such as mean and maximum) of all input features F, and inputs this global context information into a small network to dynamically generate a set of hyperedge prototypes. Each prototype can be regarded as an abstract higher-order semantic concept learned from data (for example, "objects with reflective surfaces", "backgrounds with complex textures", etc.). Then, through the multi-head attention mechanism, the module calculates the similarity between each sample feature f_j and each hyperedge prototype, thereby obtaining a "soft" association matrix that defines the extent to which each sample participates in each higher-order semantic concept.

2) Hypergraph Convolution: Next, through a two-stage message passing process, information exchange occurs between nodes (samples) and hyperedges (higher-order concepts). Firstly, the node features belonging to the same hyperedge are aggregated to form the feature representation of the hyperedge. Then, the features of these hyperedges are propagated back to the associated nodes. After this process, the initial feature f_j of each sample absorbs the context information of the high-order semantic cluster it belongs to, and finally outputs a refined high-order feature f'_j.

Ultimately, we obtain the refined high-order feature set F' = f'_1,... f'_M. Compared with F, each feature vector in F' contains information about its position and role in the entire dataset topology. After obtaining the refined high-order feature set F', we comprehensively evaluate the value of each sample x_j in the candidate pool C_i through two core indicators.

Representation Shift: We believe that after a sample is refined through the global context, the degree of change in its feature vector directly reflects its importance in the overall topological structure of the data. A structurally critical sample will have a strong high-order association with many other samples, resulting in a significant correction of its features during the extraction process. However, for a redundant or isolated sample, the feature changes will be very small. Therefore, we define the higher representative score of the JTH sample as the L2 norm between the eigenvectors before and after its refinement, that is, the "representation change" $S_{shift}(j)$:

$$S_{\text{shift}}(j) = \|\mathbf{f}'_j - \mathbf{f}_j\|_2 \tag{1}$$

where $\mathbf{f}_j \in \mathbb{R}^D$ is the initial feature vector of sample x_j, and $\mathbf{f}'_j \in \mathbb{R}^D$ is the high-order feature vector refined by the adaptive hypergraph convolution network $\mathcal{H}$, i.e., $\mathbf{f}'_j = \mathcal{H}(\mathbf{F})_j$. $\|\cdot\|_2$ denotes the Euclidean norm.

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

Prediction Confidence: Unlike the traditional active learning approach of prioritizing "uncertain" samples, our experiments show that choosing samples that the model can already recognize with high confidence is a more effective strategy. As a powerful quality filter, the high-confidence c_j ensures that the core semantics of the selected enhanced samples remain clear and unambiguous even after drastic changes. Its definition is as follows:

$$c_j = \max(\mathbf{p}_j) \tag{2}$$

where $\mathbf{p}_j$ is the predictive probability distribution vector for sample x_j from model M_i.

Combined Score: To integrate the structural importance of a sample with the reliability of the model's prediction for it, we first normalize both scores across all M candidate samples using Z-score standardization to eliminate scale differences:

$$\hat{S}_{\text{shift}}(j) = \frac{S_{\text{shift}}(j) - \mu(S_{\text{shift}})}{\sigma(S_{\text{shift}})} \tag{3}$$

$$\hat{c}_j = \frac{c_j - \mu(c)}{\sigma(c)} \tag{4}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ represent the mean and standard deviation over the entire candidate pool, respectively. The final combined score for the j-th sample, $\text{Score}(j)$, is then defined as the sum of these two standardized scores:

$$\text{Score}(j) = \hat{S}_{\text{shift}}(j) + \hat{c}_j \tag{5}$$

We select the top-n samples with the highest scores, considering them as "high-quality anchor" samples that can most effectively help the model consolidate its core knowledge and stabilize its decision boundaries for the next stage of training.

4. Experiment

In this section, we conducted a comprehensive experimental verification of the proposed PADA framework and Hyper-Ada screening strategy. Our aim is to answer the following core questions:

1) Can our Active Data augmentation framework (PADA) improve model performance compared with the standard supervised learning baseline?

2) Is the Hyper-Ada screening strategy based on high-order features that we proposed superior to the traditional screening strategy based on "uncertainty + clustering"?

3) How do these methods perform in datasets of different complexities and scenarios with varying degrees of data scarcity?

4) Does our method have an advantage in terms of computational cost?

4.1 Experimental Setup

Datasets: We conducted experiments on three image classification benchmark datasets of different complexities: CIFAR-10, CIFAR-100[20], and SVHN [21]. To simulate scenarios with scarce data, we set different subset-ratios for each dataset, namely 100% (1.0), 50% (0.5), and 20% (0.2).

Model and training details: Our model architecture selection refers to the standard practices in the fields of semi-supervised and self-supervised learning to ensure fair comparisons on a strong and recognized baseline. Specifically, we use the WideResNet-28x2[22] model on CIFAR-10, and the larger-capacity WideResNet-28x8[22] model on the more complex CIFAR-100 and SVHN. Although these Settings were borrowed, all of our experiments were conducted in a fully supervised mode. All experiments were trained end-to-end using the SGD optimizer, with a total of 200 epochs of training rounds. For the PADA framework, we set the active filtering period K (Active-Interval) to 10 epochs. The number of enhanced samples n selected each time is fixed at 10% of the size of the base training set (A_{sub}) used.

Comparison method:

1) Baseline: The standard supervised learning method. The model is trained only on the corresponding proportion of the training set (A_{sub}), without any active screening.

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

2) PADA (Entropy+Diversity): Our PADA framework, but it adopts the traditional "entropy uncertainty +K-Means clustering diversity" as the screening strategy.

3) Hyper-Ada: Our PADA framework adopts the ultimately proposed screening strategy based on adaptive hypergraphs and "determinism + representation of change".

4.2 Main Experimental Results and Analysis

Table 1 summarizes the final test accuracy rates of the main experiments. From the data, we can draw the following key observations and conclusions:

Table 1: Final test accuracy rates (%) of Different methods under various datasets and different training ratios (ρ).

Dataset	Model	K	ρ	q	baseline acc	PADA acc	Hyper-Ada acc
cifar10	WideResNet 28×2	10	0.2	0.1	88.11	88.44	88.53
			0.5	0.1	92.89	92.89	92.97
			1.0	0.1	95.01		95.34
cifar100	WideResNet 28×8	10	0.2	0.1	62.33	62.39	62.87
			0.5	0.1	74.14	74.19	74.27
			1.0	0.1	80.55		81.03
svhn	WideResNet 28×8	10	0.2	0.1	95.69	95.80	95.89
			0.5	0.1	96.91	96.94	96.95
			1.0	0.1	97.49		97.51

1) The universal effectiveness of the Hyper-Ada strategy: In all comparative experiments, the Hyper-Ada method we proposed achieved the best performance. This strongly demonstrates the superiority and universality of our screening strategy based on high-order feature extraction.

2) The sparser the data, the more obvious the advantage: The performance improvement (ΔAcc) of Hyper-Ada over the Baseline is most significant when the data is sparse. For example, on the CIFAR-100 dataset, when 100% data was used, Hyper-Ada (81.03%) improved by 0.48% compared to Baseline (80.55%); When the data volume was reduced to 20%, this performance improvement expanded to 0.54% (62.87% vs 62.33%). This validates our hypothesis: when the available labeled data is limited, it becomes particularly crucial to intelligently mine information from "cost-free" augmented data.

3) The more complex the task, the more obvious the advantages: The advantages of Hyper-Ada are more prominent on more challenging datasets. On the CIFAR-100 with more categories (20% data), the performance improvement was 0.54%, while on the relatively simple CIFAR-10 (20% data), the performance improvement was 0.42%. This indicates that our method can effectively assist the model in learning more complex feature distributions, which is crucial for distinguishing fine-grained categories.

4) Hyper-Ada outperforms the traditional PADA strategy: By comparing the results of Hyper-Ada and PADA (based on entropy + clustering), we can clearly see the superiority of the screening strategy based on high-order features. For example, on CIFAR-100 (50% data), the performance of Hyper-Ada (74.27%) is significantly higher than that of traditional PADA (74.19%). This proves that the "certainty + representation change" index we proposed can more accurately locate the most valuable enhanced samples for the model compared to the traditional "uncertainty + cluster diversity".

4.3 Ablation Experiment: The influence of the enhanced sample retention ratio

Table 2 lists the influence of different proportions of enhanced samples on the model performance. Under the setting of a 20% training data ratio, we explored the impact of increasing the number of retained "high-quality" enhanced samples after each active screening on model performance. The experimental results show that when the retention ratio increases from 10% to 50%, the final accuracy of the model shows a slight but continuous improvement (for example, on CIFAR-10, it increases from 88.53% to 88.84%). However, the marginal benefit of such performance improvement is diminishing. This indicates that although increasing the number of high-quality enhanced samples is beneficial, the significance of the screening process itself may be even higher. Even if only the top 10% of the samples are retained, significant performance improvements can already be achieved, demonstrating that our PADA framework strikes a good balance between computational efficiency and performance gain.

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

Table 2: The test accuracy of enhanced samples in different proportions(q).

Dataset	Model	K	ρ	q	baseline acc	Hyper-Ada acc
cifar10	WideResNet 28×2	10	0.2	0.1	88.11	88.53
			0.2	0.2	88.11	88.52
			0.2	0.4	88.11	88.58
			0.2	0.5	88.11	88.84
cifar100	WideResNet 28×8	10	0.2	0.1	62.33	62.87
			0.2	0.2	62.33	62.65
			0.2	0.4	62.33	62.92
			0.2	0.5	62.33	62.81
svhn	WideResNet 28×8	10	0.2	0.1	95.69	95.89
			0.2	0.2	95.69	95.88
			0.2	0.4	95.69	95.86
			0.2	0.5	95.69	95.88

4.4 Computational Cost Analysis

In addition to model accuracy, the PADA framework we proposed also demonstrates significant advantages in computational efficiency. Compared with a "brute-force augmenting" baseline - that is, pre-generating massive (for example, 10 times) augmenting data and conducting the entire training process - our method significantly reduces the training cost through a dynamic loop of "screening - training".

Take the use of 20% data (with a A_{sub} size of 10,000) on CIFAR-10 as an example. The training set size of the brute-enhanced baseline will reach 100,000, while the training set size of Hyper-Ada at each stage is only 11,000. This means that, under the same batch_size, the number of batches that the brute-force enhancement method needs to process per epoch is nearly nine times that of Hyper-Ada. Although our method introduces periodic active filtering overhead (approximately equivalent to the time consumption of 1-2 training epochs), its total training time is significantly lower than that of brute-force enhancement methods. Our approach replaces the huge "ineffective computational cost" of continuous training on a large amount of redundant data with a smaller, one-off "intelligent decision-making cost". In addition, PADA is a storage and memory-efficient strategy that does not require caching massive amounts of augmented data on disk.

4.5 Discussion on Framework Advantages

Our PADA framework and Hyper-Ada strategy also have two important conceptual advantages:

1) Decoupling and plug-and-play of augmentation strategies: The PADA framework decouples the "generation" and "use" of data augmentation, enabling it to be flexibly combined with any existing or future data augmentation techniques (such as RandAugment, CutMix, etc.) as a meta-strategy. The Hyper-Ada filter plays a role in quality control, which enables us to confidently use more aggressive and diverse enhancement methods to generate candidate pools without worrying about introducing excessive noise, greatly enhancing the universality and robustness of the framework.

2) Provide model interpretability and insights: Our approach is not merely a performance improvement tool, but also a framework for model diagnosis and analysis. The subset of samples selected by Hyper-Ada in each screening stage can be regarded as an "intelligent detection" of the current learning state of the model. By conducting visualization and statistical analysis on these selected samples, we can gain valuable insights into the learning bottlenecks of the model and the effectiveness of data augmentation strategies, providing a valuable by-product for the study of model interpretability.

5. Conclusion

In this paper, we propose a novel framework called Progressive Active Data Augmentation (PADA) to address the blindness and inefficiency issues existing in data augmentation in deep learning. This framework innovatively applies the "intelligent screening" concept of active learning to a "cost-free" data augmentation process. By periodically and adaptively optimizing the training set, it transforms data augmentation from a static and random process into a dynamic and intelligent closed loop.

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

As the core of this framework, we further designed and implemented an advanced screening strategy - Hyper-Ada. The core breakthrough of this strategy lies in that it no longer assays each enhanced sample in isolation. Instead, inspired by the idea of adaptive hypergraph networks, it takes the features of all candidate samples as a global context and uses adaptive hypergraph convolution (AdaHGConv) to extract features that contain high-order semantic associations among samples. Based on this, we have proposed a brand-new sample value assessment system, which combines the "certainty" of model prediction with the "representation change" we define. Experiments have proved that screening for samples of "high-quality anchor points" that the model can confidently identify and also play a key role in the overall data structure is a more effective strategy than traditional methods based on uncertainty.

We conducted comprehensive comparative experiments on three benchmark datasets (CIFAR-10, SVHN, CIFAR-100). The results clearly show that:

1) The Hyper-Ada strategy we proposed consistently outperforms the standard supervised baseline and the active screening method based on traditional clustering in all experimental setups.

2) The advantages of this method are particularly prominent in scenarios with sparse data (for example, using only 20% of the training data) and complex tasks (such as CIFAR-100), demonstrating its great potential in solving problems of limited data in the real world.

3) Compared with brute force enhancement methods, our framework significantly reduces the computational and storage resources required for training while achieving better performance, realizing a win-win situation of efficiency and effectiveness. In conclusion, our work provides an effective and efficient new paradigm for how to intelligently utilize data augmentation. By introducing an understanding of the higher-order structure of data into active learning, we have demonstrated that the shift from "choosing what to label" to "choosing what to learn" is feasible and productive.

For future work, we believe there are several directions worth exploring. Firstly, more complex multi-scale feature fusion modules (such as HyperACE) can be integrated into our high-order feature extraction process in order to capture richer semantic information. Secondly, the PADA framework can be applied to a wider range of task domains, such as text enhancement in natural language processing or lesion recognition in medical image analysis. Finally, through a more in-depth analysis of the samples selected by Hyper-Ada, it is expected to provide valuable insights for the interpretability of the model and the design of a better data augmentation strategy itself.

References

[1] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (NIPS), 25 (2012).
[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016).
[3] Lewis, D.D., Gale, W.A.: A sequential algorithm for training text classifiers. In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3–12 (1994).
[4] Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In International Conference on Learning Representations (ICLR) (2018).
[5] Lei, M., Li, S., Wu, Y., et al.: YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception. arXiv preprint arXiv:2506.17733 (2025).
[6] DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017).
[7] Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, 34(07), 13001–13008 (2020).
[8] Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In International Conference on Learning Representations (ICLR) (2018).
[9] Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: CutMix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6023–6032 (2019).
[10] Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 113–123 (2019).
[11] Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In Proceedings of the IEEE/CVF Conference on Computer Vision and

ISSN 2616-5775 Vol. 8, Issue 11: 52-61, DOI: 10.25236/AJCIS.2025.081106

Pattern Recognition Workshops (CVPRW), pp. 702–703 (2020).
[12] Settles, B., Craven, M., Ray, S.: Multiple-instance active learning. In Advances in Neural Information Processing Systems (NIPS), 21 (2008).
[13] Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep batch active learning by diverse gradient embeddings. In International Conference on Learning Representations (ICLR) (2020).
[14] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations (ICLR) (2017).
[15] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. In International Conference on Learning Representations (ICLR) (2018).
[16] Berge, C.: Graphs and Hypergraphs. North-Holland Publishing Company (1973).
[17] Zhou, D., Huang, J., Schölkopf, B.: Learning with hypergraphs: Clustering, classification, and embedding. In Advances in Neural Information Processing Systems (NIPS), 19 (2006).
[18] Gao, Y.W., Wang, J., Zhang, Z., Ji, R., Wu, Y.: Hypergraph-induced semantic tuplet loss for deep metric learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7646–7655 (2022).
[19] An, C., Chen, S., Zhang, Z., Wang, X., Zhang, Z.: Graph-based active learning for memory-efficient deep learning. arXiv preprint arXiv:2106.14227 (2021).
[20] Krizhevsky, A.: Learning multiple layers of features from tiny images. Technical Report, University of Toronto (2009).
[21] Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning. In NIPS Workshop on Deep Learning and Unsupervised Feature Learning (2011).
[22] Zagoruyko, S., Komodakis, N.: Wide residual networks. In British Machine Vision Conference (BMVC) (2016).