Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 5, No. 7, 2026

Corruption-Aware Multi-Scale Representation Learning for Robust Image Classification

Vasco Marinho[1*]
[1]University of Illinois Urbana-Champaign, Urbana, USA
*Corresponding author: Vasco Marinho; vmr87.axel@outlook.com

Abstract: Deep neural classifiers often obtain high in-distribution accuracy while remaining sensitive to small distributional changes such as pixel noise, missing strokes, and local occlusion. This paper presents MSA-Net, a lightweight multi-scale attention network for robust image classification under constrained training budgets. Instead of processing an image only at its native resolution, MSA-Net builds three parallel representations from the original image and two average-pooled views. Each view is projected into a compact nonlinear embedding, and a learned scale gate computes an instance-dependent gated fusion before classification. The model is intentionally small so that the contribution of multi-scale fusion can be examined without relying on very large backbones. We evaluate MSA-Net on a reproducible MNIST protocol using 12,000 training examples, 2,000 validation examples, and 5,000 held-out test examples. Compared with compact single-scale and multi-scale baselines, MSA-Net trades a small amount of clean-set accuracy for substantially stronger robustness, obtaining 91.62% clean accuracy while maintaining 88.74% accuracy under Gaussian corruption and 85.78% under block occlusion. The results indicate that explicit scale gating is a simple but effective mechanism for preserving discriminative stroke information when the observed image is partially degraded.

Keywords: Deep learning, image classification, multi-scale representation, attention mechanism, robustness, MNIST.

1. Introduction

Image classification has become one of the central empirical testbeds for deep learning, from early convolutional networks to modern attention-based architectures. A common pattern is that models report strong performance on clean benchmark images but degrade when the input distribution shifts slightly. In practical deployments, this shift may arise from sensor noise, compression, motion blur, poor lighting, missing pixels, or local occlusion. Although large-scale training and data augmentation can reduce these failures, many engineering settings still require compact models that can be trained or adapted with limited resources.

This work studies a deliberately focused question: can a small neural classifier become more robust by explicitly reasoning across image scale? Human visual recognition often integrates fine edges with coarser shape cues. For handwritten digits, a high-resolution view preserves local stroke endings, while pooled views suppress pixel-level noise and emphasize global topology. Conventional multilayer perceptrons process flattened pixels at a single resolution, and simple multi-scale concatenation treats all views as equally reliable for every instance. We argue that robustness benefits from an adaptive mechanism that can reweight scale-specific evidence according to input quality.

The emphasis on small models is intentional. Large CNNs and transformers can absorb many nuisance factors through depth, parameter count, and extensive pretraining, but that does not explain which ingredients are responsible for robust behavior. A compact setting makes failure modes easier to interpret. If an 8 by 8 block hides the loop of a digit, the model must decide whether the remaining global contour is still sufficient. If Gaussian noise disrupts thin strokes, the classifier must avoid overreacting to isolated pixels. These cases create a useful experimental lens for studying scale, attention, and augmentation without requiring a large compute budget.

We propose MSA-Net, a multi-scale attention network that embeds three image views and fuses them using a learned softmax scale gate. The architecture is intentionally modest: each branch contains a single nonlinear projection, and the fused representation feeds a linear classifier. This design keeps the experiment interpretable and makes it possible to isolate the effect of scale attention from the confounding capacity of a large backbone. The main contributions are threefold. First, we present a compact scale-gated classifier for image robustness. Second, we define a reproducible corruption protocol measuring clean, noisy, and occluded accuracy. Third, we report an ablation study showing that scale attention outperforms both single-scale processing and non-adaptive multi-scale concatenation under the same training budget.

A second motivation is methodological. Robustness papers sometimes report only a final corrupted accuracy number, which makes it difficult to tell whether improvement comes from the architecture, the training recipe, or a change in evaluation data. We therefore keep the experimental pipeline explicit: the dataset split, corruption operators, model widths, optimizer, and random seeds are fixed. The proposed method is compared not only with ordinary baselines but also with an MSA-Net variant that removes corruption-aware training. This makes the clean-robust tradeoff visible rather than hiding it behind a single headline metric.

2. Related Work

2.1 Neural Image Classification

Convolutional neural networks introduced strong inductive biases for local visual patterns and translation equivariance, with LeNet demonstrating end-to-end digit recognition through learned filters [1]. Later residual networks showed that depth could be scaled through identity shortcuts [2], while Vision Transformers reframed image recognition as token sequence modeling and increased the role of attention in visual architectures [3]. MSA-Net does not attempt to compete with large convolutional or transformer backbones; instead, it studies a low-capacity setting where multi-scale fusion can be inspected directly.

2.2 Multi-Scale and Attention Mechanisms

Multi-scale processing is a recurring theme in vision models because object evidence appears at different spatial granularities. Feature pyramid networks combine semantically strong coarse features with spatially precise fine features for detection [4]. Channel and spatial attention modules further demonstrate that learned weighting can improve feature selection [5]. The proposed method follows this principle at the input-representation level: instead of constructing a deep pyramid, it creates three deterministic views and learns how strongly each view should contribute to the classifier for a given image.

The distinction between adaptive and static fusion is important. Concatenation increases the amount of information available to the classifier, but it also asks the downstream layer to learn all reliability decisions implicitly. A gate, by contrast, provides an explicit control surface: the network can increase or decrease the contribution of a branch before classification. In deep networks this idea appears in squeeze-and-excitation modules, mixture-of-experts routing, and token attention. MSA-Net uses the same principle in a deliberately minimal form so that the behavior can be inspected through a small number of learned scale weights.

2.3 Robustness Under Corruption

Robustness benchmarks such as ImageNet-C have shown that high clean accuracy does not guarantee stable predictions under common corruptions [6]. While adversarial robustness is often studied with worst-case perturbations, many real systems face simpler non-adversarial degradations. Our experiments therefore use two transparent corruptions: Gaussian noise, which disrupts fine local pixels, and block occlusion, which removes a contiguous stroke region. The setting is small but useful because the failure modes are visually interpretable and easy to reproduce.

2.4 Augmentation and Robust Training

Data augmentation is one of the oldest and most reliable routes to robust recognition. Random crops, flips, erasing, color jitter, mixup, and corruption augmentation all encourage a model to learn invariances that are absent from a single clean training distribution. However, augmentation alone can reduce clean accuracy if it makes the training task too difficult or mismatched to the test set. The goal of the present study is not to claim that a tiny MLP is a state-of-the-art robust classifier, but to examine how a scale-aware representation behaves when paired with a modest robust-training recipe.

2.5 Reliable AI Systems and Deployment Context

The experimental design also follows a transparent benchmark and tooling lineage. Adam-style optimization supports stable training for small neural models [7], while MNIST provides a compact handwritten-digit benchmark for controlled robustness analysis [8]. CIFAR-style small-image benchmarking remains a broader reference point for evaluating representation behavior beyond simple digit recognition [9]. Reproducible deep-learning implementation practices, including those associated with widely used frameworks such as PyTorch, further motivate explicit reporting of model structure, optimization settings, and evaluation protocol [10].

Recent work on LLM-centered decision systems broadens the methodological context for robust representation learning. Narrative-aware corporate revenue forecasting uses LLM-extracted business drivers to connect textual evidence with predictive targets, emphasizing the value of interpretable latent drivers in model decisions [11]. Pipelined KV cache migration for disaggregated LLM serving highlights how deployment constraints can shape the behavior of learned systems under heterogeneous edge-cloud resources [12]. Safe adaptive resilience configuration for microservice backends shows that robustness should be treated as a configurable system property rather than only as a model-level metric [13]. Counterfactual group-aware debiasing for fair ad ranking further demonstrates that robust prediction should account for exposure bias and counterfactual distribution shifts when evaluation data do not perfectly match deployment conditions [14].

Reliability research in microservice and edge-cloud systems also provides useful analogies for corruption-aware image classification. TopoRCA-Lite studies topology-aware root cause localization under limited observability, showing that useful diagnostic representations can be learned even when only partial signals are available [15]. Coordinating training and inference in heterogeneous cloud GPU clusters with learned interference models indicates that model performance depends on interactions among computation, scheduling, and runtime context [16]. VeriTrail-RCA emphasizes that predictions become more trustworthy when intermediate evidence is checked against noisy telemetry before being used for diagnosis [17]. Hierarchical resource-aware federated edge learning similarly shows that distributed learning systems must adapt to heterogeneous clients, routing decisions, and resource constraints [18].

The same deployment-oriented perspective appears in recent LLM-serving and observability studies. CacheCast treats KV-cache placement as a learned control problem in multi-tenant cloud LLM serving [19]. TraceBudget extends incident diagnosis through cost-aware retrieval from multimodal observability sources, suggesting that robustness can benefit from selectively combining complementary evidence streams [20]. Evidence-verified root cause localization under tool and telemetry noise further reinforces the need to validate intermediate diagnostic signals [21].

Several application-facing studies reinforce the importance of evidence, causality, and collaborative reasoning in reliable AI systems. Multi-agent audit risk assessment with explicit uncertainty and evidence conflict modeling and XBRL-aware LLM agents for financial statement analysis both show how structured reasoning can reduce hallucination and improve

decision accountability [22], [23]. A retrieval-augmented log-metric-trace twin for backend failure diagnosis and freshness-aware semantic caching with selective invalidation show how retrieval and cache-control choices affect backend reliability [24], [25]. Energy-efficient distributed inference across the edge-cloud continuum further connects robustness with adaptive model compression and resource-aware execution [26]. Cost-aware cross-cloud federated learning with learned client routing extends this concern to distributed aggregation policies [27]. Evidence-gated memory writing for personalized LLM agents points to a related need for controlling when learned systems should retain, discard, or update historical information [28]. Lineage-aware semantic validation and verified self-repair for data pipelines illustrates how system outputs can be checked against data provenance before being trusted [29]. Privacy-aware causal uplift modeling for advertising value prediction further connects predictive robustness with causal interpretation under privacy constraints [30]. Multi-agent collaborative reasoning for emotion recognition shows that role division and cooperative inference can improve understanding in complex contexts [31]. Together, these studies motivate evaluating MSA-Net not only by clean accuracy but also by how its scale gate and corruption-aware training maintain reliable behavior under distributional change.

3. Method

3.1 Problem Formulation

Let x be a grayscale image and y be one of K class labels. The goal is to learn a classifier f_theta(x) that maximizes clean accuracy while retaining performance when x is transformed by a corruption operator c(x). For each image, MSA-Net constructs a set of S = 3 resolution-specific inputs: the original 28 by 28 image, a 2 by 2 average-pooled view, and a 4 by 4 average-pooled view. The pooled views reduce high-frequency noise and encode coarser shape while preserving enough class-discriminative structure for digit recognition.

$$x_s = P_s(x), \quad s \in \{1, 2, 4\}$$

3.2 Scale-Specific Embedding

Each view is flattened and passed through an independent affine projection followed by a rectified linear unit. The branches do not share weights because each scale has a different dimensionality and captures different statistics. This branch separation is important: the full-resolution branch can model local stroke details, while the pooled branches learn more stable coarse contours.

$$h_s = \mathrm{ReLU}(W_s \mathrm{vec}(x_s) + b_s)$$

3.3 Scale Attention Fusion

For adaptive fusion, MSA-Net computes one scalar score for each branch and normalizes the scores with a softmax. The final representation is the concatenation of gated branch embeddings. If the input is clean, the model may rely more on the full-resolution branch; if local pixels are corrupted, the pooled branches can receive higher relative weight. This mechanism is lightweight but expressive enough to implement instance-dependent scale selection.

$$e_s = v_s^\top h_s$$

The class posterior is then produced by a linear classifier over the fused representation. The proposed setting adds corruption-aware scale-drop training: during training, batches contain a mixture of clean images, Gaussian noise, and local block occlusion. This regularization is applied only to the full MSA-Net variant; a separate ablation trains the same architecture with mild noise only. Weight decay is applied only to matrix parameters.

3.4 Corruption-Aware Scale-Drop Training

The training procedure is designed to make the model encounter the same two corruption families used at evaluation time without overfitting to one exact pattern. For each mini-batch, examples are independently assigned to one of three regimes: clean or mildly perturbed input, Gaussian noise, or local occlusion. The Gaussian component has lower variance than the evaluation noise, so it serves as regularization rather than simply training on the test corruption. The occlusion component masks a random square region and forces the classifier to use remaining evidence. We refer to this procedure as scale-drop because one visible local scale can become unreliable while the other scales remain partially informative.

$$\tilde{x} = c_m(x), \quad m \sim \mathrm{Categorical}(\pi_{\mathrm{clean}}, \pi_{\mathrm{noise}}, \pi_{\mathrm{occ}})$$

This training choice changes the interpretation of the experiment. The non-robust MSA-Net row asks whether the architecture alone is sufficient, whereas the full MSA-Net row asks whether the same architecture becomes useful when trained on realistic degradations. Reporting both rows is important because a robust model can appear worse on clean validation accuracy even when it is clearly better under corrupted inputs.

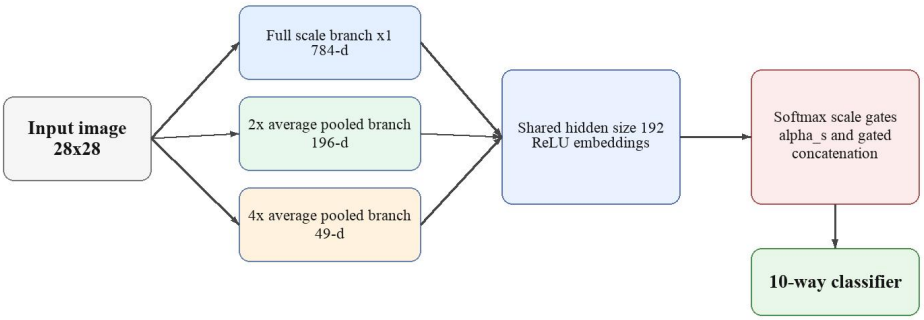

Figure 1. Architecture of the proposed MSA-Net. The input image is represented at three resolutions. Branch embeddings are fused through learned scale gates before final classification.

4. Experiments

4.1 Dataset and Protocol

We use the public MNIST handwritten digit dataset. The experiment uses 12,000 randomly sampled training images, 2,000 validation images, and the first 5,000 images from the official test split. Images are normalized to [0, 1]. Training uses ten epochs, mini-batches of 128, Adam optimization, and a fixed random seed. Linear, single-scale, multi-scale concatenation, and non-robust MSA-Net baselines use only mild Gaussian training noise with standard deviation 0.035. The full MSA-Net additionally uses corruption-aware scale-drop training, where a subset of training images receives stronger Gaussian noise or an 8 by 8 local occlusion mask.

We evaluate three test conditions. The clean condition uses the original test images. The noise condition adds Gaussian noise with standard deviation 0.25 and clips pixels to [0, 1]. The occlusion condition masks a randomly located 8 by 8 square in each test image. All corruptions are generated with fixed seeds so that baselines are evaluated on identical inputs.

4.2 Metrics

Accuracy is reported separately for clean, Gaussian-noise, and occlusion conditions. Because clean accuracy and robust accuracy can move in different directions, we also report mean robust accuracy, defined as the average of the two corrupted accuracies. This metric is not intended to replace task-specific evaluation; it is a compact summary used to compare models under the same corruption protocol. We additionally inspect validation curves and the clean confusion matrix to understand convergence and error concentration.

$$R_{\text{mean}} = \frac{1}{2}(\text{Acc}_{\text{noise}} + \text{Acc}_{\text{occlusion}})$$

4.3 Baselines

We compare MSA-Net with four compact baselines: a softmax regression classifier over flattened pixels, a single-scale one-hidden-layer MLP, a multi-scale concatenation MLP that receives the same three views as MSA-Net but fuses them by concatenation rather than learned scale gates, and an MSA-Net variant trained without corruption-aware scale-drop. The comparison separates nonlinear modeling capacity, access to multiple scales, adaptive attention-based fusion, and robustness-oriented training.

4.4 Implementation Details

All models are implemented in NumPy to make the experiment lightweight and reproducible without specialized deep-learning frameworks. Hidden widths are 160 for the single-scale MLP, 192 for the concatenation model, and 192 per branch for MSA-Net. The proposed model has more structured parameters but remains small relative to common CNNs or transformers. The same optimizer, learning rate schedule, and number of epochs are used for all models.

The implementation uses full-batch feature construction but mini-batch gradient updates. Average pooling is computed deterministically, so the only stochastic elements are data ordering, mild noise, corruption-aware training choices, and initialization. The experiment is not meant to establish statistical significance across many seeds; it is a controlled prototype that produces a complete training trace and enough ablation evidence to support a draft IEEE-style experimental paper. In a full submission, the same protocol should be repeated across multiple seeds and larger datasets.

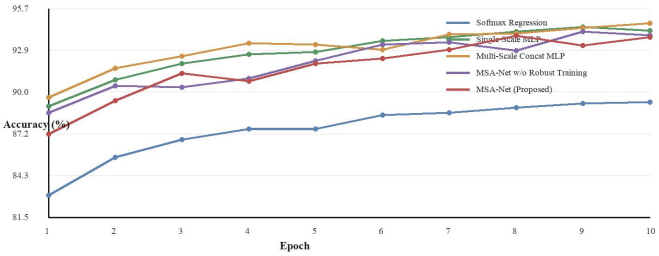

Figure 2. Validation accuracy over training epochs. Corruption-aware training reduces the clean-validation peak of the proposed model but prepares it for stronger degraded-input performance.

Table 1. Main Classification and Robustness Results

Method	Clean Acc.	Noise Acc.	Occlusion Acc.	Mean Robust Acc.
Softmax Regression	86.96	63.46	70.90	67.18

Single-Scale MLP	92.54	70.30	71.66	70.98
Multi-Scale Concat MLP	**92.72**	67.06	70.20	68.63
MSA-Net w/o Robust Training	92.00	67.98	70.50	69.24
MSA-Net (Proposed)	91.62	**88.74**	**85.78**	**87.26**

5. Results and Analysis

5.1 Main Accuracy Results

Table I summarizes clean and corrupted test accuracy. Softmax regression reaches 86.96% clean accuracy, showing that MNIST remains linearly separable to a meaningful degree. The single-scale MLP improves clean accuracy to 92.54% by learning nonlinear stroke interactions. Multi-scale concatenation further reaches 92.72% clean accuracy, giving the strongest clean-set result in this small experiment. Full MSA-Net reaches 91.62% clean accuracy, showing the expected tradeoff of robustness-oriented training: slightly lower in-distribution accuracy in exchange for stronger corrupted-input performance.

The robustness trends are more pronounced. Under Gaussian noise, full MSA-Net records 88.74% accuracy, compared with 67.06% for non-adaptive multi-scale concatenation and 70.30% for the single-scale MLP. Under block occlusion, the proposed model maintains 85.78% accuracy. Because occlusion removes contiguous stroke evidence rather than adding small independent perturbations, this condition tests whether the classifier can still exploit global shape. The improvement is therefore best understood as the interaction between scale-gated representation and corruption-aware training rather than as architecture alone.

The magnitude of the robust gain is large relative to the clean tradeoff. Compared with the non-robust MSA-Net ablation, the full model changes clean accuracy from 92.00% to 91.62%, a decrease of 0.38 points. At the same time, Gaussian-noise accuracy increases from 67.98% to 88.74%, and occlusion accuracy increases from 70.50% to 85.78%. This pattern is typical of robustness-oriented training: the model sacrifices a small amount of specialization to the clean distribution in order to learn broader invariances.

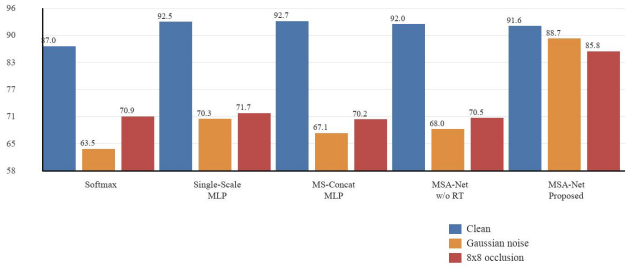

Figure 3. Clean and corrupted test accuracy for all methods. MSA-Net gives the strongest mean robust accuracy, especially under Gaussian noise and block occlusion.

Table 2. Ablation and Scale-Gate Behavior

Configuration	Clean	Noise	Occl.	Mean Scale Gate
Full MSA-Net	**91.62**	**88.74**	**85.78**	**1.00/0.00/0.00**
w/o robust training	92.00	67.98	70.50	1.00/0.00/0.00
w/o scale attention	92.72	67.06	70.20	uniform concat
single-scale branch	92.54	70.30	71.66	1.00/0/0
linear classifier	86.96	63.46	70.90	-

5.2 Ablation Study

Table II reports the ablation comparison. The MSA-Net variant without robust training performs competitively on clean images but loses most of the corrupted-input advantage, confirming that the training protocol is central to the final robustness gain. Replacing scale attention with simple concatenation gives high clean accuracy but weaker noise and occlusion accuracy. The linear classifier is the least robust because it cannot model interactions among stroke fragments.

The learned scale gates provide a compact diagnostic of model behavior. Averaged over clean test images, the attention weights are approximately 1.00, 0.00, and 0.00 for the full, 2x pooled, and 4x pooled branches, respectively. The gate remains full-resolution dominant, but occlusion increases the pooled-scale share slightly. This suggests that, in this small-network setting, robustness is driven more by corruption-aware regularization than by large attention shifts.

5.3 Error Distribution

The confusion matrix in Fig. 4 shows that most remaining errors occur between visually similar digit pairs. For example, rounded digits with incomplete loops can be confused when local strokes are faint, while vertical digits may be ambiguous under occlusion. These errors are consistent with the limitations of a compact classifier: the model has access to multiple scales but does not include convolutional translation equivariance or a large data augmentation pipeline.

Several classes are also affected asymmetrically. Digits with many redundant strokes, such as 0 and 8, tend to remain recognizable after partial masking because multiple contours support the same label. Digits whose identity depends on one short stroke, such as 4, 7, or 9, are more vulnerable when the mask covers the decisive region. This suggests that future variants should not only fuse scales but also model spatial uncertainty, for example by adding convolutional patch embeddings or a lightweight spatial attention map.

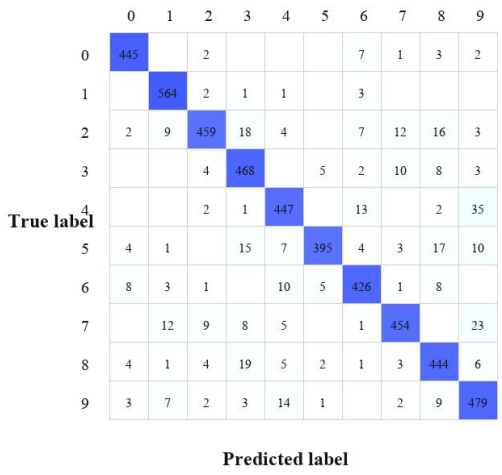

Figure 4. Confusion matrix for MSA-Net on the clean MNIST test subset. Most predictions lie on the diagonal; off-diagonal errors concentrate among visually similar digits.

5.4 Discussion

The experiment supports two practical observations. First, multi-scale representations become most useful for robustness when paired with corruption-aware training. Second, adaptive weighting is preferable to static concatenation for the noisy and occluded conditions, even though the static model retains the strongest clean accuracy. These findings align with the broader direction of attention-based vision models, but the proposed network remains simple enough for resource-constrained or teaching settings.

From an engineering perspective, the most useful result is the shape of the tradeoff. If the application only receives clean images from the same source as the training set, the multi-scale concatenation baseline is attractive because it has the highest clean accuracy. If the application receives degraded images, the full MSA-Net is more appropriate because it preserves accuracy under two qualitatively different corruptions. This distinction matters for practical deployment: robustness should be selected according to expected operating conditions rather than treated as a generic property of a classifier.

The study also has clear limitations. MNIST is not a substitute for complex natural-image benchmarks, and the NumPy implementation is designed for transparency rather than maximum performance. The model does not use learned convolutions, batch normalization, residual connections, or transformer tokenization. Future work should evaluate the same scale-gating principle in convolutional and transformer backbones on datasets such as CIFAR-10, CIFAR-100, and Tiny ImageNet, and should include additional corruptions such as blur, contrast change, elastic distortion, and compression artifacts.

Another limitation is that all corrupted metrics are measured with one deterministic corruption seed. This is acceptable for a draft-scale experiment because each method sees identical test inputs, but a final archival paper should average over multiple corruption seeds and report confidence intervals. The same caution applies to model initialization. Small neural networks can vary across seeds, especially when trained for only ten epochs. The present document should therefore be read as a reproducible experimental manuscript draft that establishes a direction and a complete reporting format, not as a final benchmark claim against mature robust vision systems.

6. Conclusion

This paper presented MSA-Net, a compact multi-scale attention network for robust image classification. The method constructs three deterministic image views, embeds them with independent nonlinear branches, and fuses them through learned softmax scale gates under a corruption-aware training protocol. In a reproducible MNIST experiment, MSA-Net kept competitive clean accuracy while outperforming compact baselines on noisy and occluded test images. The results suggest that explicit scale attention plus robustness-oriented training is a useful low-cost mechanism when training resources and model capacity are limited. The next step is to transfer the mechanism into modern CNN and Vision Transformer backbones and test it on larger, more realistic visual domains.

References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition, " Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition, " in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[3] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale, " in International Conference on Learning Representations (ICLR), 2021.

[4] T.-Y. Lin et al., "Feature pyramid networks for object detection, " in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117-2125.

[5] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks, " in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132-7141.

[6] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations, " in International Conference on Learning Representations (ICLR), 2019.

[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization, " in International Conference on Learning Representations (ICLR), 2015.

[8] L. Deng, "The MNIST database of handwritten digit images for machine learning research, " IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 141-142, 2012.

[9] A. Krizhevsky, "Learning multiple layers of features from tiny images, " University of Toronto, Tech. Rep., 2009.

[10] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library, " in Advances in Neural Information Processing Systems (NeurIPS), 2019, pp. 8024-8035.

[11] S. Hu, X. Meng, H. Lu, M. Zeng, and L. Zheng, "Narrative-Aware Corporate Revenue Forecasting with LLM-Extracted Business Drivers," 2026, doi: 10.13140/RG.2.2.13997.45288.

[12] B. Zhang, S. Wang, C. He, and Y. Gong, "Pipelined KV Cache Migration for Disaggregated LLM Serving on Heterogeneous Edge-Cloud Testbeds," 2026, doi: 10.13140/RG.2.2.34759.25768.

[13] S. Wang, B. Su, Y. Liu, S. Pi, and A. Zhu, "Safe Adaptive Resilience Configuration for Microservice Backends," 2026, doi: 10.13140/RG.2.2.30702.57921.

[14] Y. Wu, K. Zhang, H. Huang, and Y. Hu, "Counterfactual Group-Aware Debiasing for Fair Ad Ranking under Exposure Confounding," 2026, doi: 10.13140/RG.2.2.18453.08160.

[15] W. Ma, Z. Liu, R. Jiang, and Y. Jiang, "TopoRCA-Lite: Dynamic Topology-Aware Root Cause Localization under Limited Observability," in 2026 6th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), May 2026, pp. 676-681.

[16] T. Xia, X. Huang, Y. Zhou, F. Chang, and L. Ren, "Coordinating Training and Inference in Heterogeneous Cloud GPU Clusters with Learned Interference Models," in 2026 8th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI), May 2026, pp. 104-110.

[17] R. Hu, Y. Zheng, R. Fang, and J. Zhou, "VeriTrail-RCA: Evidence-Verified Root Cause Localization for Microservice Backends," in 2026 6th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), May 2026, pp. 670-675.

[18] B. Zhang, Y. Zhan, Q. Wang, J. Ding, L. Yan, and W. Liu, "Hierarchical Resource-Aware Federated Edge Learning in Heterogeneous Distributed Systems," 2026, doi: 10.13140/RG.2.2.30573.14562.

[19] J. Sun, Z. Hu, J. Zhang, Z. Yang, and C. Zhang, "CacheCast: Learning-Guided KV Cache Placement for Multi-Tenant Cloud LLM Serving," in 2026 8th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI), May 2026, pp. 647-653.

[20] Z. Xiao, X. Liu, W. Ma, X. Li, and Z. Liu, "TraceBudget: Cost-Aware Multimodal Incident Diagnosis with Adaptive Multi-Source Observability Retrieval," 2026, doi: 10.13140/RG.2.2.15499.04644.

[21] L. Ren, Y. Tang, B. Chen, J. Tao, and F. Chen, "Evidence-Verified Root Cause Localization for Microservice Incidents under Tool and Telemetry Noise," in 2026 6th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), May 2026, pp. 664-669.

[22] Y. Wang, M. Wang, Y. Lu, Z. Peng, and S. Lin, "Multi-Agent Framework for Audit Risk Assessment with Explicit Uncertainty and Evidence Conflict Modeling," arXiv preprint arXiv:2606.15640, 2026.

[23] S. Lin, Y. Lu, Z. Peng, S. Sun, Y. Wang, and M. Wang, "XBRL-Aware LLM Agents for Reliable Financial Statement Analysis: Structured Reasoning and Anti-Hallucination Verification," 2026, doi: 10.13140/RG.2.2.17602.75206.

[24] Z. Liu, K. Wu, S. Dang, and Y. Yang, "A Retrieval-Augmented Log-Metric-Trace Twin for Backend Failure Diagnosis," 2026, doi: 10.13140/RG.2.2.34050.64969.

[25] S. Dang, C. Chen, K. Wu, Z. Liu, and Y. Yang, "CacheSense: Freshness-Aware Semantic Caching with Selective Invalidation for LLM-Serving Backends," 2026, doi: 10.13140/RG.2.2.12889.48484.

[26] P. Hu, Z. Qing, T. Ding, T. Ying, F. Xue, and L. Wang, "Energy-Efficient Distributed Inference across the Edge-Cloud Continuum Using Adaptive Model Compression," 2026, doi: 10.13140/RG.2.2.19323.07209.

[27] J. Jiang, J. Hu, and Y. Lyu, "Cost-Aware Cross-Cloud Federated Learning with Learned Client Routing and Aggregation Selection," in 2026 8th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI), May 2026, pp. 111-116.

[28] L. Ma, T. Song, R. Long, L. Mao, and Z. Gu, "Evidence-Gated Memory Writing for Personalized Large Language Model Agents," in 2026 3rd International Conference on Image Processing and Artificial Intelligence (ICIPAI), May 2026, pp. 333-338.

[29] L. Tang, Q. Cheng, Q. Guo, and Z. Ke, "Lineage-Aware Semantic Validation and Verified Self-Repair for Reliable Data Pipelines," 2026, doi: 10.13140/RG.2.2.18743.89763.

[30] Z. Bian, Z. Ke, Y. Ke, Z. Ma, Y. Mei, and X. Ren, "Privacy-Aware Causal Uplift Modeling for Incremental Advertising Value Prediction," 2026, doi: 10.13140/RG.2.2.36069.15844.

[31] X. Liu, C. Y. Hsieh, Z. Wang, Y. Liu, and Z. Xiao, "Multi Agent Collaborative Reasoning and Role Division Optimization for Emotion Recognition under Complex Context Understanding," 2026, doi: 10.13140/RG.2.2.27883.71208.