

Promoting Shape Bias in CNNs: Frequency-Based and Contrastive Regularization for Corruption Robustness

Robin Narsingh Ranabhat, Longwei Wang, Amit Kumar Patel, and KC santosh

USD AI Research Lab, University of South Dakota, Vermillion SD 57069, USA

Emails: robin.ranabhat@coyotes.usd.edu, longwei.wang@usd.edu,
amit.patel@coyotes.usd.edu, kc.santosh@usd.edu

Abstract. Convolutional Neural Networks (CNNs) excel at image classification but remain vulnerable to common corruptions that humans handle with ease. A key reason for this fragility is their reliance on local texture cues rather than global object shapes—a stark contrast to human perception. To address this, we propose two complementary regularization strategies designed to encourage shape-biased representations and enhance robustness. The first introduces an auxiliary loss that enforces feature consistency between original and low-frequency filtered inputs, discouraging dependence on high-frequency textures. The second incorporates supervised contrastive learning to structure the feature space around class-consistent, shape-relevant representations. Evaluated on the CIFAR-10-C benchmark, both methods improve corruption robustness without degrading clean accuracy. Our results suggest that loss-level regularization can effectively steer CNNs toward more shape-aware, resilient representations.

Keywords: Frequency Domain Filters · Regularization for Robustness · Visible Image Corruptions · Shape vs. Texture Bias in CNNs · Contrastive learning.

1 Introduction

Convolutional Neural Networks have achieved remarkable success on benchmark datasets such as ImageNet and CIFAR-10, demonstrating high accuracy in a wide range of image classification tasks. However, their performance significantly degrades in the presence of common image corruptions, such as noise, blur, compression artifacts, and environmental distortions like fog or frost [1, 2, 13–16, 18, 22, 24–26]. This brittleness undermines the practical deployment of CNNs in real-world applications, including autonomous driving, medical imaging, and surveillance systems, where input data is rarely clean or curated. Recent work by Geirhos et al. [3] revealed that standard CNNs trained on datasets like ImageNet predominantly rely on local texture patterns rather than global object shape for making predictions. This observation contrasts sharply with the human visual

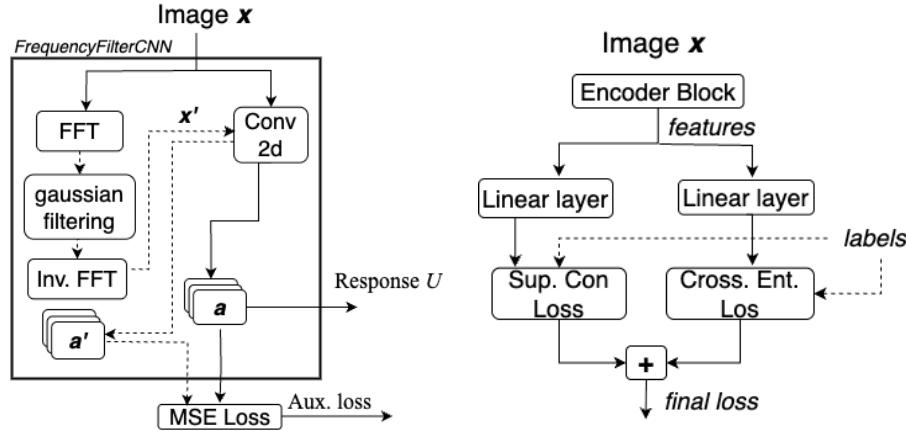


Fig. 1. Custom CNN layer which produces a separate reconstruction of original input x based on low-frequency components obtained after Fourier Transform. Activations a and a' are generated separately from the same Convolution layer. a' is cached to calculate the Mean-Squared-Loss.

Fig. 2. Integrating Supervised Contrastive Loss with Regular Cross-Entropy Loss. We hypothesize this enables learning high-level shape features existent within images of same class.

system, which is known to prioritize shape cues for object recognition. The reliance on texture renders CNNs highly susceptible to corruptions that primarily perturb high-frequency components—textures—while leaving shape information relatively intact.

Motivated by this perspective, we propose two orthogonal yet complementary regularization strategies that modify the loss function to encourage CNNs to learn representations that may exhibit reduced texture bias and enhanced shape focus. Our approach centers on the principle that appropriate regularization can guide the optimization process toward solutions that generalize better under distribution shift, particularly corruption-induced changes. Our first regularization approach operates in the frequency domain by introducing auxiliary loss terms that enforce consistency between feature activations of original images and their low-frequency filtered counterparts. This regularization mechanism biases the network toward learning features that remain stable across frequency manipulations, potentially favoring shape-relevant information that typically resides in low-frequency components over texture-specific high-frequency details. Our second regularization strategy employs supervised contrastive learning as an additional term in the training objective. Unlike traditional cross-entropy training that focuses solely on output-level classification accuracy, this regularization approach explicitly structures the intermediate feature space by pulling together representations of same-class samples while pushing apart different-class representations. We hypothesize that this representation-level regularization naturally

encourages the learning of class-invariant features, which are more likely to be shape-based than texture-based. The key insight is that strategic regularization can indirectly promote beneficial inductive biases without requiring explicit architectural changes or specialized data preprocessing. We evaluate our regularization methods on the CIFAR-10-C benchmark, which applies 19 diverse corruption types to the CIFAR-10 dataset across five severity levels, providing a rigorous testbed for corruption robustness. Our results demonstrate that both regularization strategies improve robustness while maintaining clean accuracy, showing complementary improvements across different corruption types.

2 Related Work

2.1 Robustness and Regularization in CNNs

CNNs struggle under distributional shifts caused by common image corruptions. To assess this vulnerability, Hendrycks and Dietterich [2] introduced the ImageNet-C and CIFAR-10-C benchmarks, applying real-world corruptions like Gaussian noise and motion blur. Their results showed significant performance drops even in state-of-the-art models.

In response, numerous regularization strategies have been proposed to enhance model robustness. Classical regularization techniques like dropout [3] and batch normalization [4] have been shown to improve generalization, but their specific impact on corruption robustness has been less systematically studied. Data augmentation remains a widely used approach, wherein training data is augmented with synthetic corruptions to increase diversity and reduce overfitting. Notably, methods such as AugMix [5] have shown promise in this regard. However, these techniques often require careful design and may struggle to generalize across all corruption types. Alternative solutions include neuroscience-inspired architectural enhancements, such as divisive normalization layers [7] and push-pull inhibition mechanisms [6, 8, 9, 27–30, 32, 34, 35], which aim to emulate biological visual processing.

2.2 Shape vs. Texture Bias

Geirhos et al. [10] demonstrated that conventional CNNs are strongly biased toward texture-based features rather than shape-based cues. This conclusion was drawn using Stylized ImageNet (SIN), a modified version of ImageNet in which textures were randomized via style transfer while preserving object shapes. CNNs trained on SIN showed increased shape bias and improved robustness to corruptions. This observation stands in contrast to human visual perception, which primarily relies on global shape for object recognition. The gap between CNNs and human perception has motivated follow-up studies exploring techniques to enhance shape bias in models. Some approaches include designing networks with greater receptive fields to better capture global structures [11, 36], implementing advanced data augmentation techniques that preserve structural information [12] and adopting vision transformer architectures, which inherently prioritize global shape due to their patch-based attention mechanisms [17].

2.3 Fourier Perspective on Robustness

Another lens through which CNN robustness has been analyzed is the frequency domain. Yin et al. [19] showed that CNNs tend to rely heavily on high-frequency components, which are particularly sensitive to noise and artifacts. This reliance renders models fragile under conditions that perturb fine-grained details. Building on this insight, Wang et al. [20] proposed training strategies that encourage models to focus on low-frequency components—typically associated with shape—thereby improving generalization and corruption robustness. These findings support the hypothesis that promoting low-frequency, shape-relevant representations can serve as a foundation for robust visual learning.

2.4 Contrastive Learning

Contrastive learning has recently gained prominence as a powerful framework for representation learning, particularly in the context of self-supervised learning [21]. It operates by structuring the latent space such that similar samples (positive pairs) are drawn closer together while dissimilar ones (negative pairs) are pushed apart. This principle has been extended to supervised learning by Khosla et al. [23], who demonstrated that incorporating label information into the contrastive objective significantly improves the quality of learned representations compared to standard cross-entropy training. While contrastive learning has predominantly been applied to improve general feature quality and transferability, its potential to enhance model robustness—particularly through promoting shape-based invariance—remains underexplored. Our work builds upon this emerging direction by employing supervised contrastive loss to foster representations that are both discriminative and stable across corruptions, with an emphasis on capturing class-specific shape features.

3 Proposed Methodology

We propose two complementary regularization approaches to improve CNN robustness to common corruptions by modifying the training objective through auxiliary loss terms.

3.1 Frequency-Based Regularization with Filtered Convolution

Our first approach is motivated by the observation that texture information typically resides in high-frequency components of images, while shape information is predominantly captured by low-frequency components. We introduce a novel convolution layer, *FrequencyFilterCNN*, that encourages the network to produce similar activations for both original inputs and their high-frequency filtered counterpart as shown in Fig. 1. The *FrequencyFilterCNN* operates as follows:

1. For an input tensor x , compute standard convolution activations $a_x = \text{Conv}(x)$.

2. Transform x into the frequency domain using the 2D Fast Fourier Transform (FFT): $X = \mathcal{F}(x)$.
3. Apply a Gaussian low-pass filter in the frequency domain to obtain X' , suppressing high-frequency components: $G(u, v) = \exp(-(u - u_c)^2 - (v - v_c)^2 / 2\sigma^2)$, where (u_c, v_c) is the center of the frequency domain, and σ controls the cutoff frequency.
4. Transform back (inverse FFT) to the spatial domain: $x' = \mathcal{F}^{-1}(X')$. x' also represents a high-frequency filtered version of original input x .
5. Compute activations for the filtered input: $a_{x'} = \text{Conv}(x')$.
6. During training, minimize the mean squared error between a_x and $a_{x'}$ as an auxiliary loss: $\mathcal{L}_{\text{aux}} = \text{MSE}(a_x, a_{x'})$

The final loss function combines the standard cross-entropy loss with this auxiliary loss: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda \cdot \mathcal{L}_{\text{aux}}$, where λ is a hyperparameter controlling the contribution of the auxiliary loss. By encouraging similar activations for both original and high-frequency filtered inputs, the network is biased toward features that are preserved under blurring, which we hypothesize to be more shape-related rather than texture-related. This approach differs from simply training over additional blurred images, as we have control during backpropagation to update weights of specified CNN layers. Ideally, we only swap the first few CNN layers with *FrequencyFilterCNN*, while using standard CNN for rest of downstream layers, assuming, the top layers will extract discriminative features from noisy inputs [8] [9].

3.2 Supervised Contrastive Learning

Our second approach leverages supervised contrastive learning to encourage the network to learn representations that capture class-invariant features, which are more likely to be shape-related than texture-related. As visualized in Fig. 2, we modify a standard ResNet architecture to produce two representations from the penultimate layer: $f(x)$ used for contrastive learning and $g(x)$ used for classification with cross-entropy loss. We adopt the supervised contrastive loss from Khosla et al. [23]:

$$\mathcal{L}_{\text{SupCon}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(f_i \cdot f_p / \tau)}{\sum_{a \in A(i)} \exp(f_i \cdot f_a / \tau)} \quad (1)$$

where I is the set of indices in the batch, $P(i)$ is the set of indices of examples with the same class as example i , $A(i) = I \setminus i$, and τ is a temperature parameter. The representations f are normalized to the unit hypersphere. Similar to earlier, the final cross-entropy loss is regularized using this supervised contrastive loss as: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \alpha \cdot \mathcal{L}_{\text{SupCon}}$, where α is a hyperparameter controlling the contribution of the contrastive loss. Since shape is more consistent within classes than texture (which can vary considerably even within different instances of same class), we consider this approach to naturally promote the shape bias.

4 Experimental Details

4.1 Datasets and Model Design

We conducted our experiments using the CIFAR-10 dataset [31] for training and the CIFAR-10-C benchmark [2] for evaluating robustness. CIFAR-10 consists of 50,000 training images and 10,000 test images across 10 classes, with each image being 32×32 pixels. CIFAR-10-C applies 19 different corruption types at 5 severity levels to the CIFAR-10 test set, resulting in 95 test scenarios (19 corruptions \times 5 levels). We applied minimal data augmentation during training, consisting only of random cropping and horizontal flipping. For both methods, we used ResNet-18 [33] as our backbone architecture, with following modifications specific to each method:

Low-Frequency Activation Learning We replaced the first three convolutional layers of ResNet-18 with our custom *FrequencyFilterCNN*. The *FrequencyFilterCNN* was implemented as described in Section 3.1, with the Gaussian filter’s sigma parameter set to 0.1.

During training, we computed the auxiliary MSE loss across for each *FrequencyFilterCNN* module in the network between their a_x and $a_{x'}$ activations and summed them to get final MSE loss. This loss was scaled by a factor $\lambda = 0.2$ and added to the cross-entropy loss.

Supervised Contrastive Learning For the supervised contrastive learning approach, we kept the standard ResNet-18 architecture but added two separate linear heads after the final global average pooling layer. First is a projection head for contrastive learning, consisting of a linear layer followed by L2 normalization. Second one is a classification head for standard classification, consisting of a single linear layer. The projection head mapped the flattened ResNet features to a 64-dimensional embedding space, where the contrastive loss was computed. The classification head mapped the encoded features to the probability distributions over the 10 output classes.

5 Results and Analysis

5.1 Overall Performance

Table 1 compares the performance of our proposed methods against the baseline ResNet-18 (regular CNN layers) on both clean CIFAR-10 test data and corrupted CIFAR-10-C data. Both of our proposed methods show improved robustness on CIFAR-10-C compared to the baseline. Method 2 (Supervised Contrastive Learning) achieves the best performance on both clean and corrupted data, with a 0.8% improvement in mean corruption robustness accuracy over the baseline. Method 1 (Low-Frequency Activation Learning) shows a slight decrease in clean accuracy (-0.6%) but with improvement in robustness accuracy (+0.3%).

Table 1. Accuracy (%) of different models on clean CIFAR-10 and various corruption types from CIFAR-10-C. For readability purpose, the best results are in **bold** and the second best results are underlined.

Corruption Type	ResNet-18 Baseline	Frequency Domain Reg. (Ours)	Supervised Contrastive Reg. (Ours)	Push-Pull Baseline
Clean CIFAR-10	84.82	84.23	86.19	82.63
Mean Corruption Acc	68.33	68.60	<u>69.13</u>	70.31
<i>Gaussian Noise</i>	<u>45.07</u>	44.77	42.59	62.90
<i>Impulse Noise</i>	53.44	45.91	<u>54.58</u>	62.92
<i>Shot Noise</i>	52.90	<u>53.02</u>	52.39	67.82
<i>Speckle Noise</i>	54.72	54.79	<u>54.97</u>	68.33
Avg. Noise Acc.	<u>51.53</u>	49.62	51.13	65.5
<i>Gaussian Blur</i>	69.32	71.50	<u>69.92</u>	69.03
<i>Defocus Blur</i>	74.32	75.92	<u>75.44</u>	73.53
<i>Glass Blur</i>	71.30	74.02	71.32	<u>72.53</u>
<i>Motion Blur</i>	68.98	70.06	69.84	68.61
<i>Zoom Blur</i>	70.52	72.60	<u>72.17</u>	70.86
Avg. Blur Acc.	70.89	72.82	<u>71.74</u>	70.91
<i>Brightness</i>	<u>80.98</u>	80.90	82.93	78.89
<i>Contrast</i>	<u>53.79</u>	52.55	54.56	49.74
<i>Elastic Transform</i>	77.06	<u>77.63</u>	77.93	75.40
<i>Fog</i>	<u>70.78</u>	68.84	71.37	64.88
<i>Frost</i>	71.45	74.58	<u>72.59</u>	70.45
<i>JPEG Compression</i>	74.66	75.23	<u>76.37</u>	76.79
<i>Pixelate</i>	<u>81.58</u>	82.44	81.45	80.45
<i>Saturate</i>	<u>74.80</u>	74.74	77.50	73.45
<i>Snow</i>	74.09	75.20	<u>75.13</u>	72.25
<i>Spatter</i>	78.60	<u>78.78</u>	80.47	77.11

Supervised Contrastive Learning demonstrates balanced improvements across different corruption types, although it's performance is slightly lower than the baseline on noise corruptions (-0.40% on average across all noises). To add an alternative perspective, we also present the results from Push-Pull CNN, that regularizes through architectural changes in CNN layer. It shows an exceptional performance specifically on noise corruptions (+13.97% on average). But on closer inspection, our methods still demonstrate a rather balanced performance improvement over a greater number of corruptions. These results suggest that the two methods have complementary strengths, with frequency-transform based approach particularly effective against blur corruptions, and contrastive-loss approach being performant against more numbers of individual corruptions.

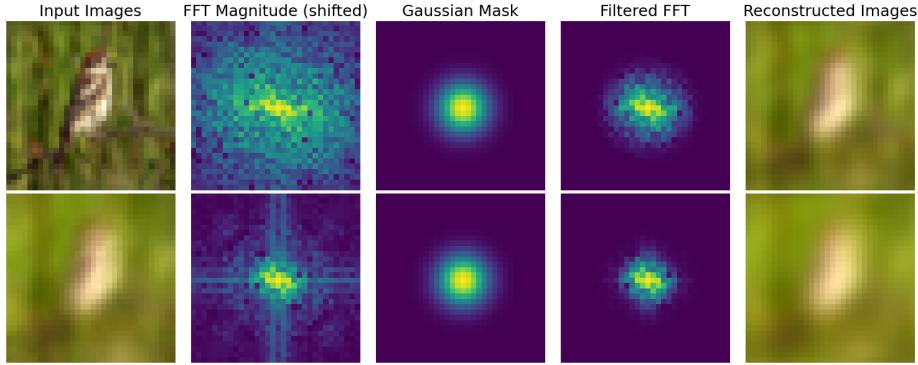


Fig. 3. Visualization of *FrequencyFilterCNN* processing on clean and corrupted images. The figure illustrates the key processing steps: original input images (first column), Gaussian mask applied in frequency domain (second column), FFT magnitude spectrum (third column), filtered frequency spectrum (fourth column), and resulting reconstructed images (fifth column). The top row shows processing of a clean bird image, while the bottom row demonstrates the same processing applied to a bird image corrupted with defocus blur at severity level 1.

5.2 Analysis of the Methods

Low-Frequency Reconstruction Gaussian filtering at Frequency-domain reconstructs the original signal from low-frequency components only. From Fig. 3, we can see that, applying gaussian-mask on the Frequency Domain in input image, actually has effect of Blurring. By training the network to produce similar activations for both original and their blurred counterpart, CNN learns robust kernels that are less prone to be activated by high-frequency patterns. Pixelate is another corruption where this approach benefits substantially. From Fig. 4, we find *FrequencyFilterCNN* layer to demonstrate more robust activation patterns that are less sensitive to pixelate corruption compared to the response maps of regular convolutional layer. We can see clearer feature representations maintained even at higher corruption severity. The slight decrease in performance on noise corruptions (-1.91%) suggests that this method may overly suppress high-frequency details that could be useful for distinguishing between certain classes under noise conditions. However, the overall corruption accuracy improves by 0.27% compared to the baseline, indicating that the benefits outweigh this limitation.

Supervised Contrastive Learning The regularization of cross-entropy loss with a supervised contrastive learning encourages the network to learn “patterns that statistically prominent within different instances of a class”. For example, while cats may differ in breed and color, their facial structures tend to be similar — a property that distinguishes them clearly from other classes like dogs or inanimate objects such as tables. We assume such features to be more shape-

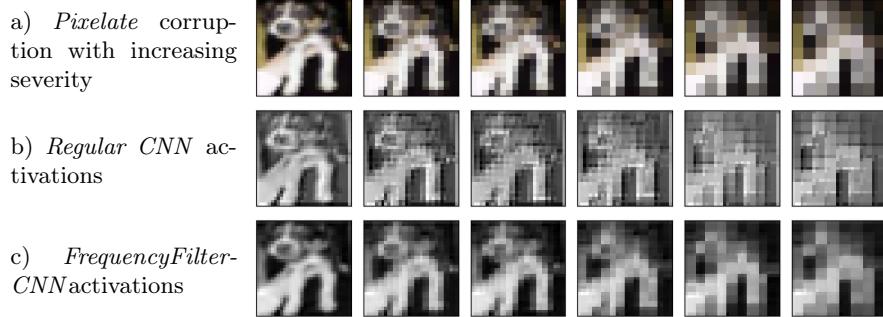


Fig. 4. Comparison of activation patterns under pixelate corruption. (a) Original CIFAR-10 dog image (leftmost) and the same image corrupted with pixelate noise at increasing severity levels 1-5 (left to right). (b) Feature activations from a regular convolutional layer when processing the images shown in (a). (c) Feature activations from the proposed *FrequencyFilterCNN* layer on the same input images.

based than texture-based with shape being considered a more reliable cue for object identity. However, there is also the possibility that the network exploits unintended correlations to minimize the loss. For example, it might latch onto background elements or dataset-specific artifacts that co-occur frequently within certain classes, rather than the truly semantic object features. Our current work does not explicitly investigate the extent to which such shortcuts are being exploited.

6 Conclusion and Future Work

In this paper, we explored two complementary regularization approaches to enhance CNN robustness that we hypothesize should discourage texture bias while promoting shape bias: a custom CNN architecture that enforces similarity between activations from original and low-frequency reconstructed input, and a supervised contrastive learning approach that should promote shape-focused representations. Our results demonstrate that both methods effectively improve robustness against common corruptions on the CIFAR-10-C benchmark, with the supervised contrastive learning approach showing the strongest generic performance. In future work, we will evaluate our models on stylized datasets to quantify shape and texture bias more directly. We also plan to combine both regularization methods to exploit their complementary strengths.

References

1. L. Wang, I. I. Uddin, X. Qin, Y. Zhou, and K. C. Santosh, “Explainability-Driven Defense: Grad-CAM-Guided Model Refinement Against Adversarial Threats,” in *Proceedings of the AAAI Symposium Series*, vol. 6, no. 1, pp. 49–57, 2025.

2. D. Hendrycks and T. Dietterich, “Benchmarking neural network robustness to common corruptions and perturbations,” in *International Conference on Learning Representations (ICLR)*, 2019.
3. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
4. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning (ICML)*, pp. 448–456, 2015.
5. D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, “Augmix: A simple data processing method to improve robustness and uncertainty,” in *International Conference on Learning Representations (ICLR)*, 2020.
6. L. Wang, X. Chen, Z. Zhao, and H. Zhang, “Exploration vs exploitation for distributed channel access in cognitive radio networks: A multi-user case study,” in *Proc. 2011 11th International Symposium on Communications and Information Technologies (ISCIT)*, pp. 360–365, 2011.
7. A. Cirincione, R. Verrier, A. Bic, S. Olaiya, J. J. DiCarlo, L. Udeigwe, and T. Marques, “Implementing Divisive Normalization in CNNs Improves Robustness to Common Image Corruptions,” in *SVRHM Workshop @ NeurIPS*, 2022.
8. N. Strisciuglio, M. Lopez-Antequera, and N. Petkov, “Enhanced robustness of convolutional networks with a push–pull inhibition layer,” *Neural Computing and Applications*, vol. 32, no. 24, pp. 17957–17971, 2020.
9. R. Ranabhat, L. Wang, X. Qin, Y. Zhou, and K. C. Santosh, “Multi-Scale Unrectified Push-Pull with Channel Attention for Enhanced Corruption Robustness,” in *AAAI 2025 Summer Symposium Series (SuSS-25)*, 2025.
10. R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, “ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness,” in *International Conference on Learning Representations (ICLR)*, 2018.
11. X. Ding, X. Zhang, J. Han, and G. Ding, “Scaling up your kernels to 31x31: Revisiting large kernel design in CNNs,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11963–11975, 2022.
12. S. Lee, I. Hwang, G. C. Kang, and B.-T. Zhang, “Improving robustness to texture bias via shape-focused augmentation,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4323–4331, 2022.
13. C. Wall, L. Wang, R. Rizk, and K. C. Santosh, “Winsor-CAM: Human-Tunable Visual Explanations from Deep Networks via Layer-Wise Winsorization,” *arXiv preprint arXiv:2507.10846*, 2025.
14. L. Wang, N. Nayyem, and A. Rakin, “Enhancing Adversarial Robustness of Deep Neural Networks Through Supervised Contrastive Learning,” *arXiv preprint arXiv:2412.19747*, 2024.
15. N. Nayyem, A. Rakin, and L. Wang, “Bridging Interpretability and Robustness Using LIME-Guided Model Refinement,” *arXiv preprint arXiv:2412.18952*, 2024.
16. L. Wang, X. Li, and Z. Zhang, “Dense cross-connected ensemble convolutional neural networks for enhanced model robustness,” *arXiv preprint arXiv:2412.07022*, 2024.
17. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.

18. L. Wang, A. Ghimire, and K. C. Santosh, “Enhanced Model Robustness by Integrated Local and Global Processing,” in *2024 IEEE 6th International Conference on Cognitive Machine Intelligence (CogMI)*, pp. 234–239, 2024.
19. D. Yin, R. G. Lopes, J. Shlens, E. D. Cubuk, and J. Gilmer, “A fourier perspective on model robustness in computer vision,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.
20. H. Wang, X. Wu, Z. Huang, and E. P. Xing, “High-frequency component helps explain the generalization of convolutional neural networks,” in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8684–8694, 2020.
21. T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International Conference on Machine Learning (ICML)*, pp. 1597–1607, 2020.
22. L. Wang, P. Chen, C. Wang, and R. Wang, “Layer-wise entropy analysis and visualization of neurons activation,” in *International Conference on Communications and Networking in China (ChinaCom)*, pp. 29–36, 2019.
23. P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, “Supervised contrastive learning,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 18661–18673, 2020.
24. L. Wang, A. Ghimire, K. C. Santosh, Z. Zhang, and X. Li, “Enhanced robustness by symmetry enforcement,” in *IEEE Conference on Artificial Intelligence (CAI)*, 2024.
25. K. Xiao, L. Wang, A. Gupta, and X. Qin, “Looking Beyond Content: Modeling and Detection of Fake News from a Social Context Perspective,” in *HICSS*, pp. 1–10, 2022.
26. L. Wang, C. Wang, Y. Li, and R. Wang, “Improving robustness of deep neural networks via large-difference transformation,” *Neurocomputing*, vol. 450, pp. 411–419, 2021.
27. L. Wang, C. Wang, Y. Li, and R. Wang, “Explaining the behavior of neuron activations in deep neural networks,” *Ad Hoc Networks*, vol. 111, p. 102346, 2021.
28. L. Wang and Q. Liang, “Representation learning and nature encoded fusion for heterogeneous sensor networks,” *IEEE Access*, vol. 7, pp. 39227–39235, 2019.
29. L. Wang and Q. Liang, “Partial interference alignment for heterogeneous cellular networks,” *IEEE Access*, vol. 6, pp. 22592–22601, 2018.
30. L. Wang and Q. Liang, “Performance analysis of cooperative multicell precoding with global CSI and local individual CSI in the large dimensional regime,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3229–3238, 2017.
31. A. Krizhevsky, “Learning multiple layers of features from tiny images,” Technical Report, University of Toronto, 2009.
32. L. Wang, W. Chen, and J. Li, “Congestion aware dynamic user association in heterogeneous cellular network: A stochastic decision approach,” in *Proc. 2014 IEEE International Conference on Communications (ICC)*, pp. 2636–2640, 2014.
33. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
34. L. Wang and Q. Liang, “Optimization for user centric massive MIMO cell free networks via large system analysis,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–1, 2016.
35. M. Shi, R. Wang, E. Liu, Z. Xu, and L. Wang, “Deep reinforcement learning based computation offloading for mobility-aware edge computing,” in *International Con-*

- ference on Communications and Networking in China (ChinaCom)*, pp. 53–65, 2019.
36. Y. Li, Y. Wang, and L. Wang, “DFT-spread based PAPR reduction of OFDM for short reach communication systems,” in *International Conference in Communications, Signal Processing, and Systems*, pp. 696–703, Springer Singapore, 2019.