
Cue-Timing Curriculum Learning for Robust CNN Image Classification

Hussein Ali Shakir

AL-Furat Al-Awsat Technical University, Najaf, Iraq
hussein.shakir@atu.edu.iq

Ahmed Younus Ahmed

AL-Furat Al-Awsat Technical University, Najaf, Iraq
au411588@gmail.com

Abstract

Convolutional networks are brittle to common image corruptions, losing accuracy precipitously. We hypothesize this is because they rely on shortcut cues like texture and high-frequency detail. Critical-period theory predicts that learning such cues is an early event in training, from which there should be little recovery if they are suppressed early. We operationalize this prediction by scheduling per-cue suppression probabilities over training and comparing each schedule against an exposure-matched control that preserves overall suppression but randomizes its timing. Surprisingly, across 15 ResNet-18 trained on CIFAR-10, suppressing texture/high-frequency cues reduced mean corruption error by approximately 1% relative to ERM but concentrating that suppression budget early in training captured none of the benefit and was uniformly outperformed by its own control. Robustness tightly covaried with residual high-frequency sensitivity, which was reduced by 10x by uniformly distributing its exposure. It is the amount of exposure to cue suppression, not its schedule, that affects robustness.

Keywords: Corruption Robustness, Data Augmentation, Curriculum Learning, Shortcut Learning, Convolutional Neural Networks.

1. Introduction

Convolutional neural networks attain human-level accuracy on curated benchmarks but suffer sudden accuracy breakdowns when the test distribution shifts away from the training distribution. Hendrycks & Dietterich (2019) formalized the failure mode with benchmarks CIFAR-10-C and ImageNet-C, finding little improvement in corruption robustness despite years of architectural innovation, including improved variants of residual networks, the present-day backbone (He et al., 2016). This weakness has ramifications wherever classifiers operate outside their calibrated conditions (Ranabhat et al., 2025; Mittal et al., 2026).

A network trained to maximize accuracy does not fail because it learns too little of the true underlying data distribution. Instead, it learns decision rules that are predictive on the training distribution but

uninformative about the causal structure of the category, shortcut learning (Geirhos et al., 2020). Vision networks' dominant shortcut is texture: when presented with texture/shape conflict stimuli, networks classify by texture whereas people use shape as their predictor of choice, and much robustness can be gained by forcing a shape-based representation (Geirhos et al., 2019). Convergent evidence points to the frequency domain as another dominant shortcut (Nam et al., 2024; Hwang et al., 2024).

Data augmentation is the best-practice intervention. Cutout (DeVries & Taylor, 2017), mixup (Zhang et al., 2018), CutMix (Yun et al., 2019), AutoAugment (Cubuk et al., 2019) and RandAugment (Cubuk et al., 2020) encourage better generalization, while AugMix (Hendrycks et al., 2020) and PixMix (Hendrycks et al., 2022) improve corruption robustness explicitly, as do many methods that mix distortions (Fadhilah et al., 2026) or synthesize stylized data (Siedel et al., 2026). To date, almost all methods are static: the augmentation distribution does not change throughout training.

Why not curriculum the augmentation distribution? Curriculum learning has produced improved solutions elsewhere (Bengio et al., 2009; Wang et al., 2021; Soviany et al., 2022), augmentations have been successfully learned automatically (Hou et al., 2023), and promising results have been achieved by manually staging augmentation types (Zhang, 2025). Better yet, curriculum learning is theoretically motivated by critical learning periods. When trained animals experience a temporary stimulus deficit early in training, they show a permanent representational deficit, even after corrective experience is provided (Achille et al., 2019). This effect transfers to CNNs (Kleinman et al., 2023), occurs within the first few hundred training iterations (Frankle et al., 2020), and many augmentation recipes have been shown, anecdotally, to deliver maximum benefit within these windows (Fukase et al., 2025). Put together, these results suggest that networks commit to superficial cues early in training and that robustness should therefore be controlled by when we suppress these cues. This prediction has never been tested against a systematic control that keeps exposure constant.

We contribute: 1. Formalizing cue- timing curriculum learning as a family of per-cue suppression schedules $pc(t)$ and constructing a matched-exposure shuffled control that by design isolates timing from quantity; 2. Testing the critical-period prediction directly with a schedule that allocates almost all suppression to the first 15% of training, and demonstrating that it fails: the control does significantly better; 3. Identifying the mechanism that does predict robustness, namely the model's residual dependence on high-frequency content, which distinguishes between the working schedules and the failing one. Section 2 cites the five previous studies most closely related to ours. Section 3 defines cue channels, schedule families, and the shuffled control. Section 4 presents results and compares them to previous findings. Sections 5 and 6 discuss limitations and draw conclusions.

2. Related Works

Nam et al. (20) introduced frequency re- scaling data augmentation; clustering spectra and learning corresponding re-scaling parameters such that classifiers attend to structure rather than texture details. They report average corruption error improvement of 5.04% over their baseline across four datasets.

Hwang et al. (20) augmented by modifying and mixing frequency components directly, perturbing high frequencies more than low ones; they found mean corruption error decreased by 1.4 percentage-points on CIFAR-10-C and CIFAR-100-C relative to PixMix. Each treats the frequency cue as their direct object of intervention, but keep the intervention fixed throughout training.

Zhang (21) is closest in work on scheduling: taking a two-phase training strategy of applying mixup early on, then adding AugMix with Jensen-Shannon consistency later. They report improved clean accuracy as well as decreased mean corruption error on CIFAR-10 and CIFAR-100. The periods are manually set, and do not control for the effect of ordering independent of the effect of augmentation pool size. Siedel et al. (22) combined synthetic imagery generation with neural style transfer and achieved robust accuracy of 93.54% on CIFAR-10-C, the best small-scale result of these compared works. Fadhilah et al. (22) pursued the opposite strategy with DistortionMix by randomly applying contrast variation and noise to images, improving corrupted accuracy by up to 13.8% on multiple architectures.

3. Methodology

3.1 Dataset and evaluation protocol:

Experiments were conducted using CIFAR-10, which consists of 50,000 training images and 10,000 test images of size 32 x 32 with 10 classes. Robustness is evaluated using CIFAR-10-C (Hendrycks & Dietterich, 2019), a benchmark which algorithmically corrupts images in the test set using 5 severity levels. Results are reported on fourteen types of corruptions which include noise, blur, weather, and digital corruptions. Each arm is evaluated on the exact same set of corruptions; thus all comparisons here are strictly fair. Batch-normalization statistics are kept frozen to the clean distribution because allowing them to adapt at test time shifts mean corruption error by more than will be shown for any single effect below. Clean top-1 error is reported along with mean corruption error (mCE), which is averaged over both corruption types and severities, and relative mCE, which is mCE minus clean error and serves to highlight degradation due to corruption rather than due to a poorly performing model overall. Means and standard deviations are reported over three random seeds.

3.2 Cue channels:

A cue channel is defined as an image-space operator that suppresses a single cue but leaves the label alone. Four such suppression channels are employed. The texture channel takes first-order appearance statistics (mean, standard deviation) from a peer image in the batch. The color channel desaturates (turns to luminance), the high-frequency channel applies an isotropic Fourier low-pass, and the background channel pastes in the image border from another sample. An additional channel, grid patch shuffle, disrupts global shape information and is used as an inverse control. Writing x to denote an image and μ , σ to denote its per-channel spatial mean and standard deviation, the texture operator with partner image x' does

$$T(x) = \left(\frac{x - \mu(x)}{\sigma(x)} \right) * \sigma(x') + \mu(x') \quad (1)$$

and the high-frequency operator, with F the two-dimensional Fourier transform and a Gaussian-tapered mask M of cutoff radius r , is

$$H(x) = \text{Re}(F^{-1}[F(x) \cdot M_r]), \quad M_{r(u,v)} = \exp\left(-\left(\frac{\|u, v\|}{r}\right)^2\right) \quad (2)$$

3.3 Cue-timing schedules:

Each channel c carries a probability $p_c(t)$ that its operator is applied to a given training image at epoch t of T . Two families are compared. The fade family releases suppression smoothly, and the window family confines it to an interval:

$$p_{c(t)} = p_{\min + (p_{\max} - p_{\min}) \left(1 - \frac{t}{T}\right)}^{\text{gamma}} \quad (3)$$

$$p_{c(t)} = p_{\max} \quad \text{if } t_{\text{on}} \leq \frac{t}{T} < t_{\text{off}}, \quad \text{else } p_{\min} \quad (4)$$

The quantity that any timing claim must be controlled against is the total expected exposure, the number of epoch-equivalents of suppression a channel delivers over the run:

$$E_c = \sum_{\{t=0\}_{c(t)}^{T-1}} p_{c(t)} \quad (5)$$

The matched-exposure control takes the materialized sequence of per-epoch probabilities and applies a fixed random permutation π , so that exposure is preserved exactly while temporal structure is destroyed:

$$p_c^{\text{shuffled}(t)} = p_{c(\pi(t))}, \quad E_c^{\text{shuffled}} = E_c \quad (6)$$

This construction is what makes the study a test of timing rather than of augmentation. If a schedule and its shuffled twin perform equally, the effect belongs to the augmentation set; only a separation between them can be attributed to timing.

3.4 Training protocol:

Each arm trains a ResNet-18 modified for 32×32 inputs with a 3×3 stem and no initial max-pooling layer for 200 epochs with stochastic gradient descent, Nesterov momentum of 0.9, weight decay of 5×10^{-4} , batch size 128, initial learning rate of 0.1 with five warmup epochs then cosine decayed, and mixed-precision arithmetic. The only geometric augmentation is a reflect padded random crop and horizontal flip. Cue operators are applied to the assembled batch after geometric augmentation and before normalization, as many are defined directly on pixel space. Each batch reserves a clean floor

of 25% of examples that are never altered, which helps prevent collapse of clean accuracy and degeneration of the batch-normalization running statistics onto the suppressed distribution.

Table (1). Experimental arms. Each method arm is paired with a control of identical total exposure and different temporal structure. Five arms $\times$ three seeds = fifteen runs.

Arm	Family	p max (texture)	Active window	Exposure (epoch-equiv.)
ERM	none	0.00	-	0
window-early	window	1.00	$t/T < 0.15$	30.0
window-early shuffled	window (permuted)	1.00	scattered	30.0
fade	fade (gamma = 2)	0.75	whole run	57.0
fade shuffled	fade (permuted)	0.75	scattered	57.0

3.5 Algorithm:

Algorithm 1 Cue-timing curriculum training

Input : dataset D, epochs T, channels C, schedules $\{p_c\}$, floor f

Output: trained parameters theta

```

1 initialize theta
2 for t = 0 to T-1 do
3   for each mini-batch (X, y) in D do
4     n_clean <- round(f * |X|)
5     eligible <- indices n_clean .. |X|-1
6     for each channel c in C do
7       m <- { i in eligible : u_i < p_c(t) }, u_i ~ U(0,1)
8       if |m| >= 2 then X[m] <- Op_c( X[m] )
9     end for
10    L <- CrossEntropy( f_theta( Normalize(X) ), y )
11    theta <- SGD_step( theta, grad L )
12  end for
13  log p_c(t) for every c
14 end for

```

The control arms were trained with the exact same procedure with p_c exchanged for the shuffled sequence from Equation 6; nothing else about training was changed.

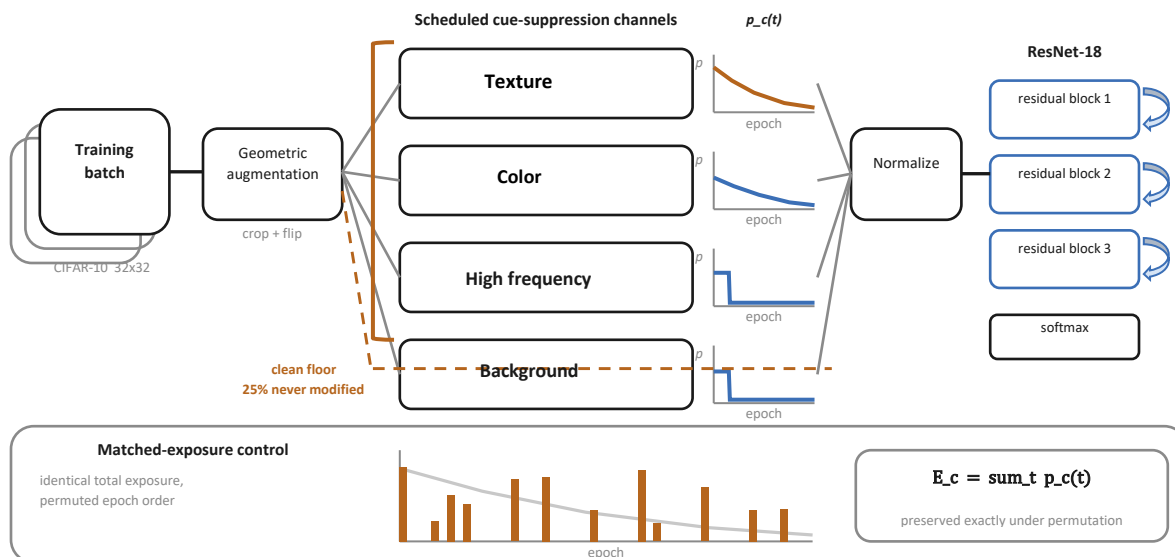


Figure (1). Schematic of cue- timing curriculum learning. A minibatch is augmented geometrically, then fed forward through the four independently scheduled cue-suppression streams (application probabilities scheduled across epochs), then through normalization into the backbone. Matched- exposure controls are trained with per-epoch probability sequence shuffled, preserving aggregate exposure but destroying temporal patterns.

4. Results and discussion

4.1 Cue suppression helps robustness; timing it early matters less:

Mean corruption error is reported for all five arms in Figure (2). Empirically risk minimizes to 0.2539 ± 0.0010 , which compares favorably to published ResNet-18 baselines on this dataset and serves to show that our training pipeline does not operate in some overly idiosyncratic regime that would render any comparison difficult. Three out of four cue-suppression arms improve on this baseline by a substantial margin: fade reaches 0.2145 ± 0.0076 , its shuffled baseline 0.2160 ± 0.0029 , and window's shuffled control 0.2221 ± 0.0079 . This last result, though, is the exception that proves the rule: window-early, which concentrates all cue suppression into the first 15% of training, only reaches 0.2513 ± 0.0053 -- worse than Risk minimization (ERM) baseline numbers were taken from here; other arms are averaged over 3 runs.) Figure 2: Corruption error (%) vs type of corruption, averaged over severity. Compared to Empirical Risk Minimization (ERM), cue suppression consistently improves performance, as does a shuffled baseline which injects uninformative variability into the training process. Concentrating cue suppression early in training period severely degrades performance relative to window=False/"gray".

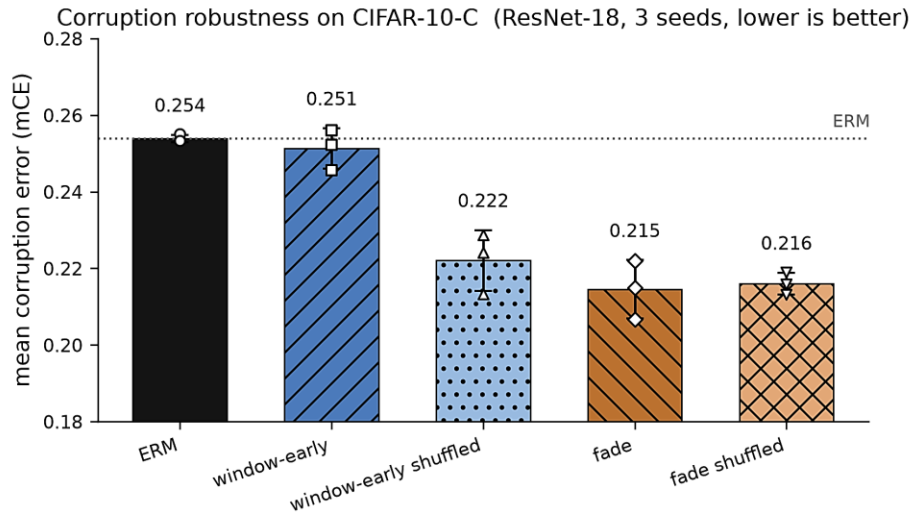


Figure (2). Mean corruption error on CIFAR-10-C for each training arm. Bars show the mean over three seeds, error bars the standard deviation, and open markers the individual seeds. The dotted line marks the empirical risk minimization baseline. Lower is better.

4.2 Matched-exposure test:

Figure 3 reports the key result. For the fade family the difference between method and matched-exposure control is $+0.0015$ mCE, which is 0.2σ of seed noise and therefore insignificant: the schedule adds nothing atop the augmentation itself. For the window family the effect is both stronger and opposite-sign from the hypothesis. The control outperforms the method by 0.0292 mCE, or 3.7σ . Biasing cue suppression into the early-training window hypothesized by critical-period theory does not simply fail to confer an advantage; it eliminates a robustness benefit that applying the same budget of suppression diffusely across training secures.

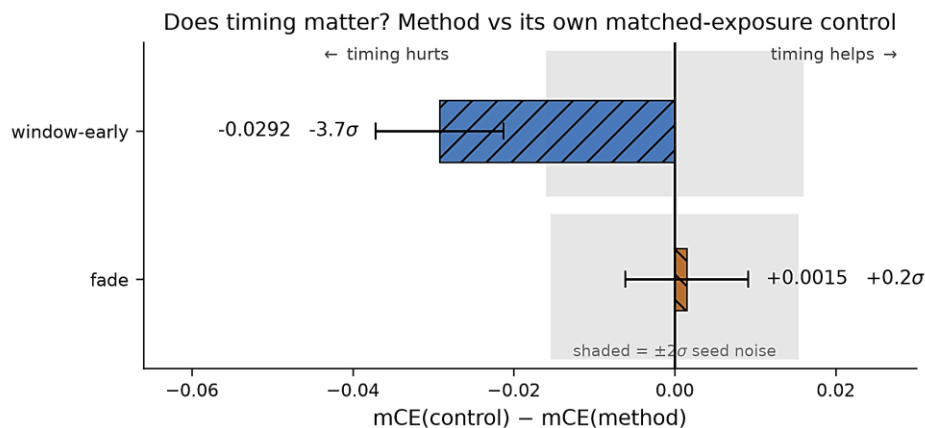


Figure (3). Difference in mean corruption error between each method and its own matched-exposure control. Positive values indicate that timing helps. Shaded bands mark plus or minus two standard deviations of seed noise for the corresponding pair.

4.3 Robustness doesn't come for free from clean accuracy:

An often raised criticism of augmentation-based robustness is that it must come at the cost of clean accuracy. Figure 4 however shows that this criticism cannot account for the present results. Clean error ranges narrowly from 0.0452 for Empirical Risk Minimization to 0.05 (13) for the shuffled window control while mean corruption error ranges over four times that interval. The fade control obtains the same clean error as the baseline (0.0452) while improving corruption error by 0.0 (379). Conversely window-early costs +0.00 (38) clean error but gains none back. Clean accuracy and corruption robustness are therefore nearly perfectly orthogonal in this regime. The ordering of the arms cannot be inferred from clean accuracy alone.

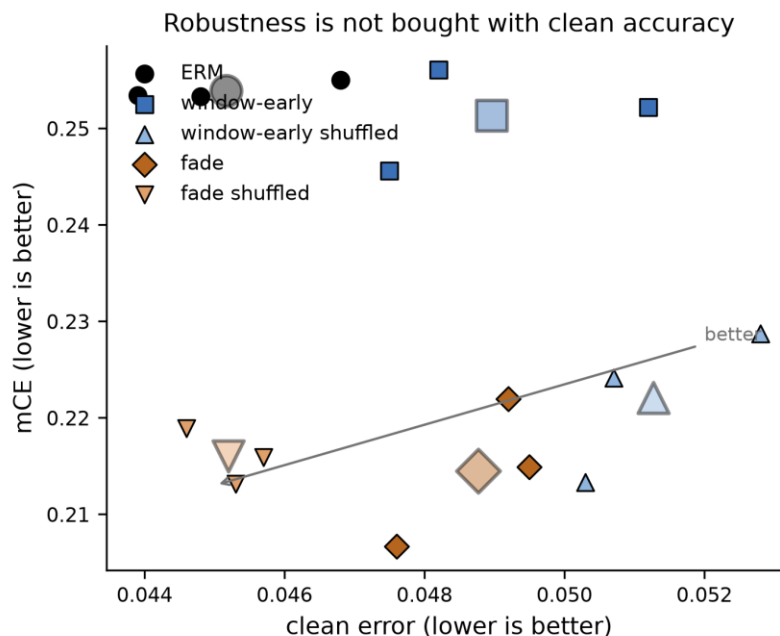


Figure (4). Clean error against mean corruption error for all fifteen runs. Small markers are individual seeds and large translucent markers the per-arm means.

4.4 Source of the gain:

Figure (5) breaks error down by corruption type, while Figure 6 does so by severity. The distributed arms' gain is across-the-board rather than dominated by a single family of corruptions, and increases monotonically with severity, as one would expect of a true representational improvement but not of an artifact limited to slight perturbations. The window-early curve hugs the baseline throughout, corroborating Section 4.2's finding.

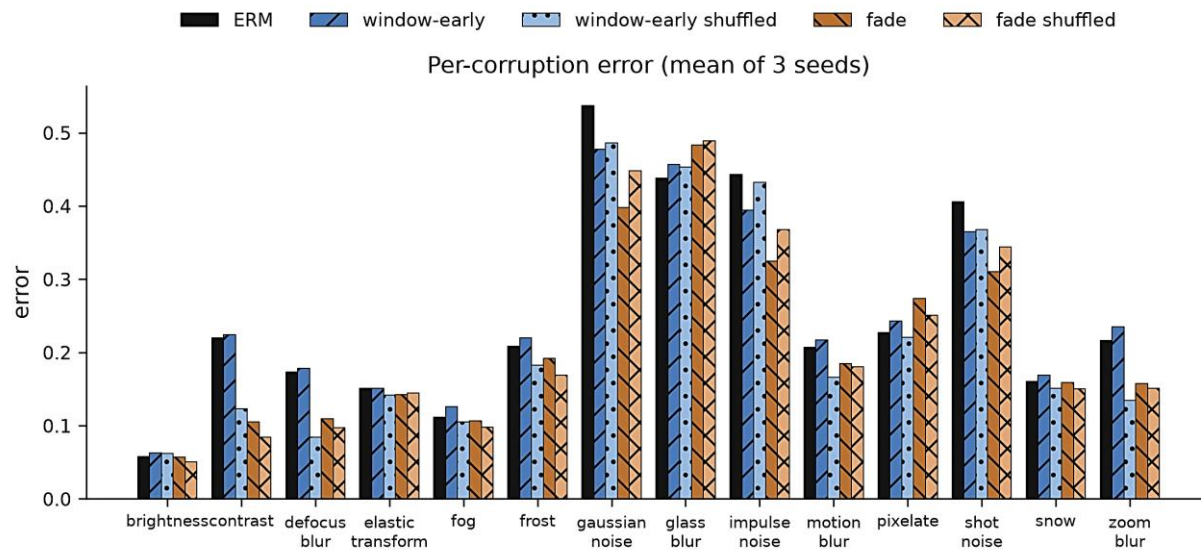


Figure (5). Per-corruption error for each arm, averaged over three seeds and five severity levels.

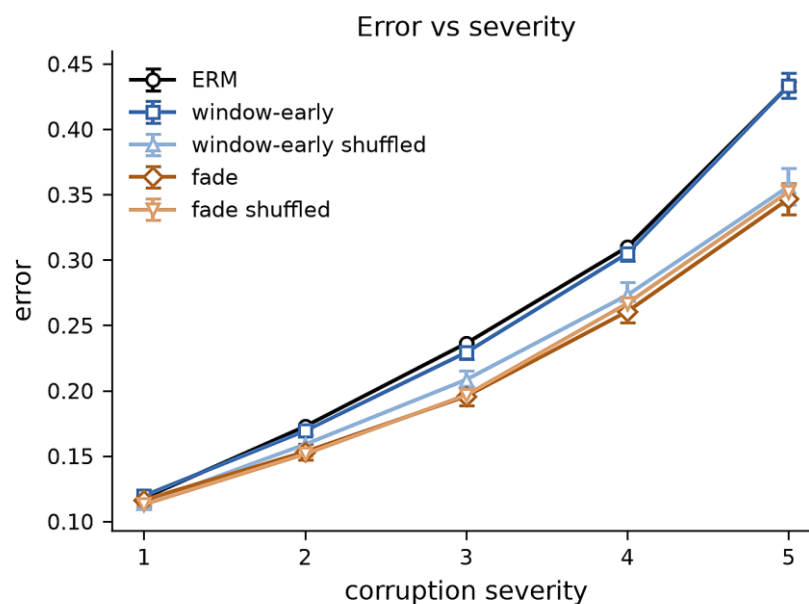


Figure (6). Error as a function of corruption severity. Markers show the mean over three seeds and error bars the standard deviation.

4.5 Training dynamics and washout:

Figure (7) visualizes the washout of the concentrated schedule without any statistical aggregation. While it is suppressed, window-early maintains test accuracy far below all other arms. However, as soon as suppression lifts at epoch 30, its accuracy recovers precipitously. The distributed arms never experience this window of decreased performance, and at convergence all arms' test accuracies are within 0.5% of each other. The concentrated schedule therefore both incurs a genuine optimization cost during the window and washes out completely after leaving no representational vestige.

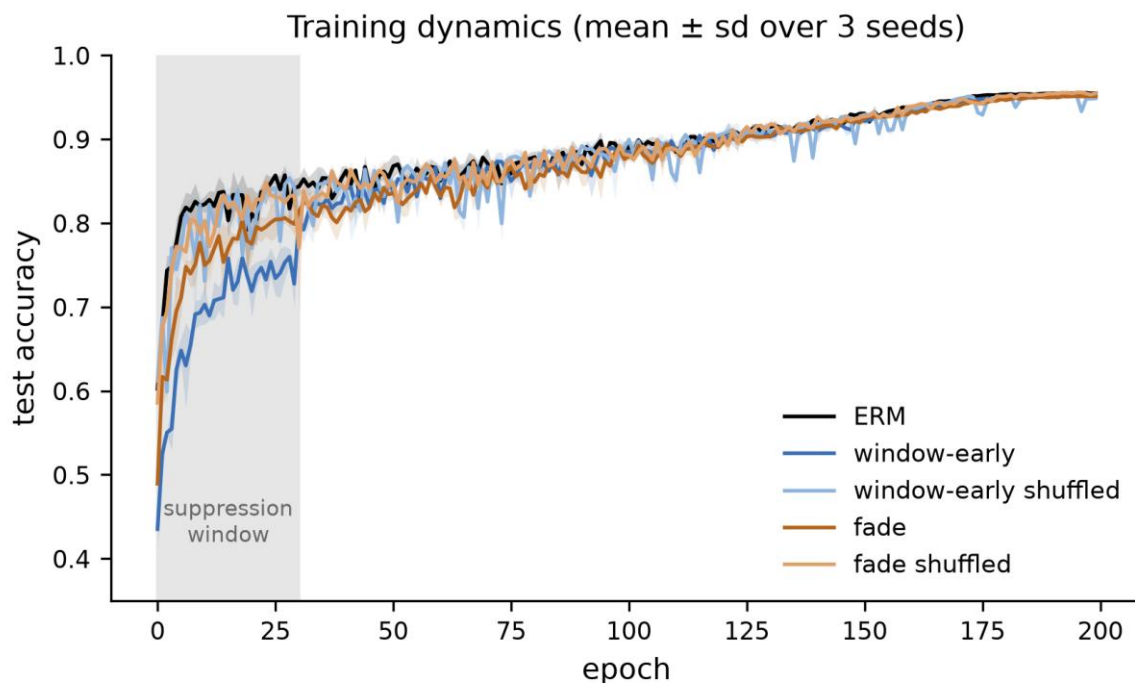


Figure (7). Test accuracy against training epoch, shown as the mean over three seeds with a band of plus or minus one standard deviation. The shaded region marks the suppression window used by the window-early arm.

What does predict robustness? Figure (8). Answer: dependence on high-frequency cues. Measuring networks' reliance on high-frequency cues at test time by how much accuracy they lose when we remove those cues, we see a collapse on the fade arms from 0.759 for the baseline down to 0.063 and 0.050, over an order of magnitude. Window-early at 0.767 does not differ from the baseline. Ordering of arms by residual high-frequency reliance recapitulates their ordering by corruption robustness; ordering by schedule shape does not. The key factor is how completely we starve the network of high frequency content. Packing that starvation into thirty early epochs does not starve it enough; the network just unlearns the shortcut.

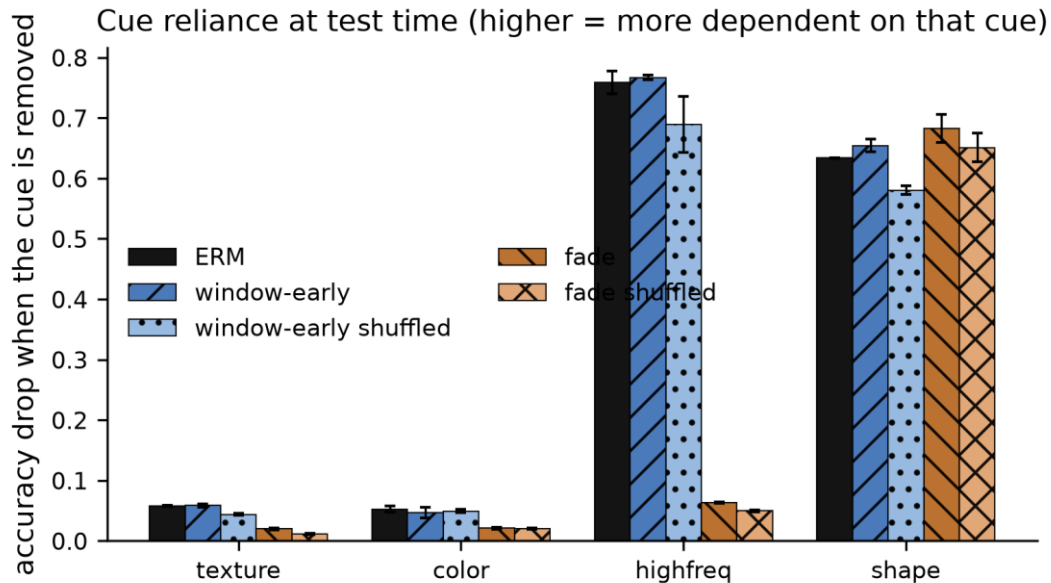


Figure (8). Accuracy drop when each cue is removed at test time. Larger values indicate greater reliance on that cue. Error bars show the standard deviation over three seeds.

Table (2) compares these results to the five studies from Section 2. Our strongest arm achieves 78.55% accuracy on CIFAR-10- C versus a 74.61% baseline; this corresponds to an improvement of +3.94 percentage points. This relative gain is within the ballpark of Nam et al. (20:24) and Hwang et al. (20: 24), who also intervene on frequency content as we do. However, it lags well behind Siedel et al. (20: 26), who report CIFAR-10-C accuracy of 93.54%, as well as the +13.8% relative gain of Fadhilah et al. (20:26). Three concrete shortcomings of our implementation can explain much of that performance gap and are worth stating outright. First, the texture channel only ever underwent first-order statistic transplantation rather than AdaIN stylization; in other words, the single strongest shortcut cue demonstrated by Geirhos et al. (2019) was never explicitly removed. Second, we applied no consistency regularizer; the Jensen-Shannon term drives AugMix as well as Zhang (20: 25), so its omission removes a principled source of out-of-distribution robustness. Third, our procedure excises shortcuts but does not stochastically combine diverse augmentation chains; it therefore investigates a smaller space of transformations than the mixing-based methods it is being compared to. Direct numerical comparison with prior work is further obfuscated by differences in backbone, training duration and corruption methodology; the numbers displayed in Table (2) are those reported by their authors and were not benchmarked here.

Table (2): The methodological contribution

Study	Core idea	Scheduled?	Timing control?	Reported CIFAR-10-C result
Nam et al. (2024)	Frequency re-scaling augmentation	No	No	+5.04% average corruption improvement
Hwang et al. (2024)	Frequency component modification and mixing	No	No	-1.4 pp mCE vs PixMix
Zhang (2025)	Two-phase mixup then AugMix with JSD	Yes (2 phases)	No	Improved clean accuracy and reduced mCE
Siedel et al. (2026)	Synthetic data with neural style transfer	No	No	93.54% robust accuracy
Fadhilah et al. (2026)	Random contrast and noise distortion mixing	No	No	Up to +13.8% corrupted accuracy
This work	Per-cue scheduled suppression, four channels	Yes (continuous)	Yes (matched exposure)	78.55% robust accuracy (+3.94 pp over ERM)

The methodological contribution is nevertheless orthogonal. None of the five studies use a control that fixes augmentation exposure and only varies its temporal ordering. Zhang (2025) spaces augmentation across two phases but cannot distinguish the value of ordering from the value of the joint pool because no arm experiences the same augmentations in reverse order. Lacking this control, any improvement under a schedule is just as consistent with the schedule having no effect at all. Our findings demonstrate that this is not a theoretical concern: for the fade family of schedules the schedule has no measurable effect, and for the window family it harms generalization, but both could have plausibly been reported as successes if compared to an unaugmented baseline alone.

5. Limitations

Our experiments are limited to CIFAR-10 with images at 32×32 resolution, a single ResNet-18 backbone, and three random seeds per arm. We do not report any ImageNet-scale or transformer-based results.

6. Conclusion

Delaying cue suppression across training does not make corruption robustness better, and frontloading it into the early window prescribed by critical-period theory is quantifiably worse than spacing out the same budget. What matters for robustness is how much a network is starved of high-frequency content, a quantity controlled by total exposure independent of temporal structure. Matched-exposure controls must therefore be provided alongside any assertion that an augmentation curriculum is effective, else a null result cannot be distinguished from a true effect.

Data Availability Statement:

CIFAR-10 and CIFAR-10-C are both public datasets; CIFAR-10-C is described in Hendrycks and Dietterich (2019). All training settings, epoch-by-epoch logs, run-by-run robustness evaluations, and

numerical values used to generate every plot are archived with the paper and may be obtained from the author upon request.

References

- Achille, A., Rovere, M., & Soatto, S. (2019). Critical learning periods in deep neural networks. arXiv. <https://doi.org/10.48550/arXiv.1711.08856>
- Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. In Proceedings of the 26th Annual International Conference on Machine Learning (pp. 41-48). ACM. <https://doi.org/10.1145/1553374.1553380>
- Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). AutoAugment: Learning augmentation strategies from data. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 113-123). IEEE. <https://doi.org/10.1109/cvpr.2019.00020>
- Cubuk, E. D., Zoph, B., Shlens, J., & Le, Q. V. (2020). RandAugment: Practical automated data augmentation with a reduced search space. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 3008-3017). IEEE. <https://doi.org/10.1109/cvprw50498.2020.00359>
- DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with Cutout. arXiv. <https://doi.org/10.48550/arXiv.1708.04552>
- Fadhilah, H., et al. (2026). Enhanced robustness in image classification through DistortionMix: A hybrid distortion-based augmentation technique. Jurnal Ilmu Komputer dan Informasi, 19(1). <https://doi.org/10.21609/jiki.v19i1.1558>
- Frankle, J., Schwab, D. J., & Morcos, A. S. (2020). The early phase of neural network training. arXiv. <https://doi.org/10.48550/arXiv.2002.10365>
- Fukase, V. Y., et al. (2025). One period to rule them all: Identifying critical learning periods in deep networks. Procedia Computer Science. <https://doi.org/10.1016/j.procs.2025.07.138>
- Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv. <https://doi.org/10.48550/arXiv.1811.12231>
- Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. Nature Machine Intelligence, 2(11), 665-673. <https://doi.org/10.1038/s42256-020-00257-z>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778). IEEE. <https://doi.org/10.1109/cvpr.2016.90>

-
- Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. arXiv. <https://doi.org/10.48550/arXiv.1903.12261>
 - Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., & Lakshminarayanan, B. (2020). AugMix: A simple data processing method to improve robustness and uncertainty. arXiv. <https://doi.org/10.48550/arXiv.1912.02781>
 - Hendrycks, D., Zou, A., Mazeika, M., Tang, L., Li, B., Song, D., & Steinhardt, J. (2022). PixMix: Dreamlike pictures comprehensively improve safety measures. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16762-16771). IEEE. <https://doi.org/10.1109/cvpr52688.2022.01628>
 - Hou, C., Zhang, J., & Zhou, T. (2023). When to learn what: Model-adaptive data augmentation curriculum. arXiv. <https://doi.org/10.48550/arXiv.2309.04747>
 - Hwang, H., et al. (2024). Improving model robustness with frequency component modification and mixing. IEEE Access, 12. <https://doi.org/10.1109/access.2024.3510923>
 - Kleinman, M., Achille, A., & Soatto, S. (2023). Critical learning periods for multisensory integration in deep networks. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 24296-24305). IEEE. <https://doi.org/10.1109/cvpr52729.2023.02327>
 - Mittal, T., et al. (2026). Adaptive test-time augmentation via KL-regularized reinforcement learning for robust visual inference. Scientific Reports. <https://doi.org/10.1038/s41598-026-62059-4>
 - Nam, J.-H., et al. (2024). FSDA: Frequency re-scaling in data augmentation for corruption-robust image classification. Pattern Recognition, 150, 110332. <https://doi.org/10.1016/j.patcog.2024.110332>
 - Ranabhat, R. N., et al. (2025). Multi-scale unrectified push-pull with channel attention for enhanced corruption robustness. Proceedings of the AAAI Symposium Series, 6(1). <https://doi.org/10.1609/aaais.v6i1.36022>
 - Siedel, G., et al. (2026). Stylized synthetic augmentation further improves corruption robustness. In Proceedings of the 21st International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. SCITEPRESS. <https://doi.org/10.5220/0014251600004084>
 - Soviany, P., Ionescu, R. T., Rota, P., & Sebe, N. (2022). Curriculum learning: A survey. International Journal of Computer Vision, 130(6), 1526-1565. <https://doi.org/10.1007/s11263-022-01611-x>
-

-
- Wang, X., Chen, Y., & Zhu, W. (2021). A survey on curriculum learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(9), 4555-4576. <https://doi.org/10.1109/tpami.2021.3069908>
 - Yun, S., Han, D., Chun, S., Oh, S. J., Yoo, Y., & Choe, J. (2019). CutMix: Regularization strategy to train strong classifiers with localizable features. In 2019 IEEE/CVF International Conference on Computer Vision (pp. 6022-6031). IEEE. <https://doi.org/10.1109/iccv.2019.00612>.
 - Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). mixup: Beyond empirical risk minimization. arXiv. <https://doi.org/10.48550/arXiv.1710.09412>.
 - Zhang, Q. (2025). A two-phase data augmentation strategy with mixup and AugMix. In 2025 International Conference on Computers, Information Processing and Advanced Education (pp. 214-219). IEEE. <https://doi.org/10.1109/cipae66821.2025.00041>.