
AuGhostmentation: The Eyes Never Stand Still—*Why Should CNNs?*

Emirhan Inan[†]
Bogazici University
Institute for DSAI
emirhan.inan814@gmail.com

Suayb S. Arslan
Bogazici University
Department of Computer Engineering
suayb.arslan@bogazici.edu.tr

Abstract

Convolutional Neural Networks (CNNs) draw broad inspiration from the primate visual system yet prove notoriously brittle under common image corruptions, whereas the latter suffers little from such degradations. Extending beyond this initial inspiration, architectures incorporating quantifiable biological priors that enhance network resilience are increasingly gaining traction. Nevertheless, most of these approaches have so far been confined to emulating the higher visual areas, seldom utilizing the upstream retinal stages that constitute the very basis of *seeing*. At its core, primate vision relies on photons captured by retinal photoreceptors, the signals of whose conversion are then relayed through the Lateral Geniculate Nucleus (LGN) to primary visual cortex (V1) and further downstream for additional processing. Because cone photoreceptors are densest at the center of the retina (fovea), the eyes must be repositioned continuously (saccade) to direct the high-acuity macula toward objects of interest (fixation). Furthermore, the resultant rapid eye movements induce intra-saccadic motion streaks that putatively aid in the spatiotemporal processing of inputs and provision of translation-invariant object correspondence across frames. Motivated by this mechanism, we propose AuGhostmentation, a portmanteau of *augmentation* and *ghosting*, which delivers the intra-saccadic retinal smear as a frontend transform. Superimposing the *ghosted* images—fainter counterparts—atop the original, based on human oculomotor statistics, the resultant datum elicits *retinal persistence* by integrating varied recompositions of the scene in a single representation, closely mimicking the foveal exploration scheme intrinsic to dynamic primate vision. When prepended to ResNet-18 trained on Tiny ImageNet, AuGhostmentation retains 98% validation accuracy while delivering a striking 60% relative gain in robustness on Tiny ImageNet-C. Applied to VOneNet, it likewise preserves 98% classification accuracy and yields a substantial 25% relative increase in compounded robustness. For ImageNet-100, the introduced *ghosts* not only improve the corruption accuracy relative to the ResNet-50 baseline by 23%, but the clean accuracy by 4%, also—with general improvements following across other varied architectures. Building on these resilience gains, AuGhostmentation outpaces competing augmentations by up to 18% and baseline ResNet-50 by up to 37% on stylized tests, suggesting that retinal priors *naturally* foster shape-biased processing and thereby provide greater alignment with human object recognition.

1 Introduction

Convolutional Neural Networks (CNNs) trace their architectural lineage to the primate visual cortex, where Hubel and Wiesel’s recordings revealed a hierarchy of cells responding to features of increasing

[†]Current Contact: www.emirhaninan.com, Code Available in: github.com/emirhaninan/AuGhostmentation

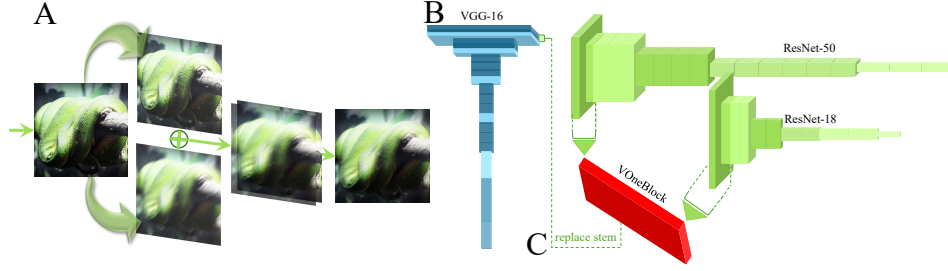


Figure 1: (A) AuGhostmentation procedure: original image and saccade-induced counterpart overlaid to produce the new input (B) VGG-16 modified to match the output of VOneNet, another biological neural network (C) Tiny ImageNet training utilizes ResNet-18 (modified to be compatible) or VOneResNet18, whereas ImageNet-100 does ResNet-50 (unmodified) or VOneResNet50. See D for the exact modification list.

complexity [Hubel and Wiesel, 1962, 1968]. Fukushima’s Neocognitron translated this biological hierarchy into a *neural network* via alternating feature-extraction and pooling layers [Fukushima, 1975, 1980], a blueprint modern CNNs extend by replacing the unsupervised learning with back-propagation trained convolutional filters, which largely deviates from the original biological motive [LeCun et al., 2015].

Despite their impressive recognition capabilities, convolutional neural networks show significant vulnerability to a varying range of image perturbations, sowing doubt in their reliability for real world applications where such distortions are especially common and possible ramifications more severe than mere misclassification [Hendrycks and Dietterich, 2019]. A natural question then arises—*what makes biological vision so resilient where neural networks fall remarkably short?* [Geirhos et al., 2018] Researchers are increasingly turning back to the source in search of answers [Sinz et al., 2019]. Biologically biased extensions of neural networks have become an emerging frontier for exploration, serving both as tools for scientific discovery and as computational models that shed light onto the mutual framework that governs visual processing across artificial and biological systems alike [Cadieu et al., 2014, Yamins and DiCarlo, 2016, Cichy et al., 2016, Richards et al., 2019, Dapello et al., 2020, Arslan, 2026].

Particularly, incorporating principles grounded in the primate visual system has been shown to increase the overall robustness of deep neural networks [Dapello et al., 2020, Baidya et al., 2021, Evans et al., 2022, Piper et al., 2025]. For instance, a well-known model VOneNet, which appends fixed Gabor filters at the fore of otherwise standard CNNs, shows substantial improvement in resilience to adversarial attacks and image corruptions [Dapello et al., 2020]. However, such apparent gains in robustness often entail sundry trade-offs—perhaps most notably, improvements in corruption accuracy usually being entailed by considerable decreases in clean image accuracy [Baidya et al., 2021, Piper et al., 2024].

This vulnerability underlines the insufficiency of hitherto emulated biological inductive biases and the need for the pursuit of newer, more robust priors [Machiraju et al., 2021]. Nevertheless, we suggest that one need not look so far downstream the cortex to find such utile bio-mimetic inspiration. Indeed, while substantial progress has been made to replicate and simulate the extensive backend systems that modulate downstream visual and cortical processing [Yamins and DiCarlo, 2016], the same precision is wanting when it comes to modeling the beginning of all vision—the *retina*.

2 Biological Background

2.1 Retina and Rapid Eye Movements

The photons that strike the eyes are not processed in stationary snapshots but in continuous streams of highly structured data. To sample its environment dynamically, primate vision relies on saccadic eye movements—rapid shifts of gaze that help foveate, that is, center objects of interest onto high-acuity regions of the retina [Kowler, 2011, Hoffman et al., 2013].

Consequently, when a target enters the periphery of the visual field, an orienting saccade is first initiated, but, to home in precisely, eyes trigger a cascade of saccades that range from broad, ambient sweeps of the scene to shorter, deliberate fixations that accumulate relevant features and aid in the construction of semantic representations [Stearns and Poletti, 2026]. Microsaccades sit at the finest scale of this cascade—long thought to merely counteract image fading but now known to actively play a role in reshaping the retinal feed through characteristic amplitude and temporal statistics [Rolfs, 2009, Engbert and Kliegl, 2003].

Unlike large saccades that reorient gaze between scenes, microsaccades occur during the fixation itself [Rolfs, 2009, Martinez-Conde et al., 2004, 2013]—a condition, we argue, more closely resembles how a static vision model processes images. Their small amplitudes ($< 1^\circ$), corresponding to only a few pixels, converted at standard field-of-view scales, make them particularly suitable for pixel grids: large enough to perturb the input meaningfully yet small enough to preserve recognition-relevant spatial content. We also note that microsaccade statistics are well characterized in the oculomotor literature [Engbert and Kliegl, 2003, Rolfs, 2009]—including lognormal amplitudes, cardinal directional bias and Poisson temporal occurrence rate—thus, altogether, enabling calibrated instead of heuristic parameterization.

2.2 Intra-saccadic Motion Smear

Pre-, intra- and post-saccadic periods have all been subjects of extensive study in vision science. The residual consequences of the first and the last are especially difficult to model in a purely feedforward network, since faithfully capturing their geometry hinges on the implementation of downstream recurrent and extraretinal cortical processes, such as remapping, corollary discharge or prediction [Melcher and Colby, 2008, Sommer and Wurtz, 2008, Wurtz et al., 2011]. The intrasaccadic period, however, is more directly relevant to model-agnostic augmentation design because it deals with the retinal consequences of the saccade itself—*what happens to the image while the eye is moving*.

The underlying motion that defines saccadic eye movements, no matter what their size, does not manifest in smooth, seamless glides. On the contrary, each saccade sweeps the retinal input afresh across photoreceptors, producing a characteristic motion smear [Bedell et al., 2004]. The brain employs several mechanisms to suppress this smear and prevent from being construed as actual translation, a phenomenon known as *saccadic retinal masking* [Duyck et al., 2016, Crevecoeur and Kording, 2017].

Saccadic masking has long been theorized to render the obfuscated, blurry signal completely invisible and irrelevant for downstream visual processing [Balsdon et al., 2018]. More recent studies, however, paint a less absolute picture: Rather than being entirely discarded, the residual smear may still be exploited as a source of information in its own right and for sustaining the spatiotemporal integration without interruption [Schweitzer and Rolfs, 2021]. In any case, the layers that succeed the retinal photoreceptors, Lateral Geniculate Nucleus (LGN) and V1, do receive signals contaminated with said blurs, however suppressed they may be [Sylvester et al., 2005, Sylvester and Rees, 2006].

2.3 Temporal Masking

Although the exact mechanism by which this suppression takes place remains contested, it is directly relevant to how an augmentation aiming to emulate it should be designed. A long-standing *active suppression* theory holds that saccade initiation is accompanied by an active extra-retinal signal that attenuates visual sensitivity to the transient smear [Matin, 1974, Volkmann, 1986, Diamond et al., 2000, Ross et al., 2001, Binda and Morrone, 2018]. The proponents of *temporal filling-in* [Campbell and Wurtz, 1978, Castet et al., 2002], on the other hand, suggest that instead of a complete neural top-down correction, what occludes the intra-saccadic retinal smear from perception is the existence of pre- and post-saccadic images in so close a temporal proximity that they absorb the transitory smear [Duyck et al., 2016]. In our implementation, likewise, each `AuGhostmentation` operates sequentially on the closely changing retinal stages. The current, sharp retinal state is taken as the pre-saccadic image, its directional blur as the intra-saccadic smear and their weighted interpolation as the post-saccadic residue—a state that is neither fully sharp nor fully smeared but carries the *ghost* of the movement—and it is this ghost, reconstructed from oculomotor data rather than suppressed, that gives `AuGhostmentation` both its mechanism and its name.

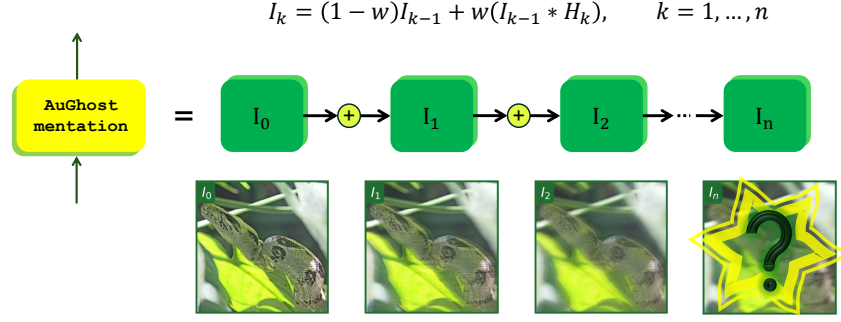


Figure 2: **The computational model of AuGhostmentation.** Each augmentation step blurs the *current composite* and folds it back in: $I_k = (1 - w)I_{k-1} + w(I_{k-1} * H_k)$, $k = 1, \dots, n$, where H_k is a microsaccade-derived blur kernel and w controls blending strength. Because each step operates on I_{k-1} rather than the original I_0 , blur accumulates non-linearly: Later microsaccades smear an already-smear image, progressively compounding the spatial perturbation.

Additionally, when multiple microsaccades occur within a period of fixation, each event takes as its input a retinal image that carries the output of its predecessor, even if ever faintly, reflecting the fact that the neural aftermath of a microsaccade persists well into the inter-saccadic period [Otero-Millan et al., 2008, Hafed et al., 2015]. This accumulated degradation incidentally helps disrupt high spatial frequencies and nudges the network towards coarser, more shape-reliant features that immanently characterize human object recognition [Landau et al., 1988, Biederman, 1987] and further from any impeding, unreliable texture shortcuts [Geirhos et al., 2019].

2.4 More than Motion

The aforementioned theories illustrate *a clear picture* of how an augmentation emulating saccades and oculomotor dynamics needs to be designed: This augmentation must have an empirically sized shift, a gradually smeared frame and an interpolation of this frame with the pre- or post-saccade image acting as masks. It must, inevitably, include intrinsic saccade characteristics such as the occurrence rate Rayner [1998], directional bias Foulsham et al. [2008] and amplitude distribution Tatler and Vincent [2008], Bahill et al. [1975].

A natural objection might be that, mechanism-wise, superimposing image copies with a streak appended on top amounts to little more than a glorified augmentation mimicking *motion blur*. Yet, as we will further demonstrate (Figure 6), this *directional blur* works most robust when oculomotor statistics are injected and the kernel size, strength, direction all biased based on biological priors.

3 Methodology

Our proposed method, AuGhostmentation, simulates each saccadic event as a biologically-bound, directional, motion-blurred kernel superimposed on the original image through a weighted blend. Four bio-mimetic parameters govern the operation, each anchored to a measurable oculomotor property: *spatial scale* via Field of View (FoV) (Section 3.1), *amplitude* and *directional bias* (Section 3.2), *occurrence rate* (Section 3.3), and *blend weight* (Section 3.4). All four were fitted from GazeBase v2.0 [Griffith et al., 2021] via the Engbert–Kliegl algorithm [Engbert and Kliegl, 2003]; derivation and cross-validation appear in Appendix C.

3.1 Field of View (FoV)

Microsaccade amplitudes are bounded by 1° ($60'$) [Zuber et al., 1965, Martinez-Conde et al., 2004], but CNNs operate in pixels. We adopt $\theta_{\text{FoV}} = 8^\circ$ following Dapello et al. [2020], who fixed this value to match studies benchmarking CNNs against primate ventral-stream recordings [Schrimpf et al., 2018, Marques et al., 2020, Cadena et al., 2019]; adopting the same FoV also ensures that the spatial scale of AuGhostmentation is commensurate with the downstream VOneNet stage. This yields

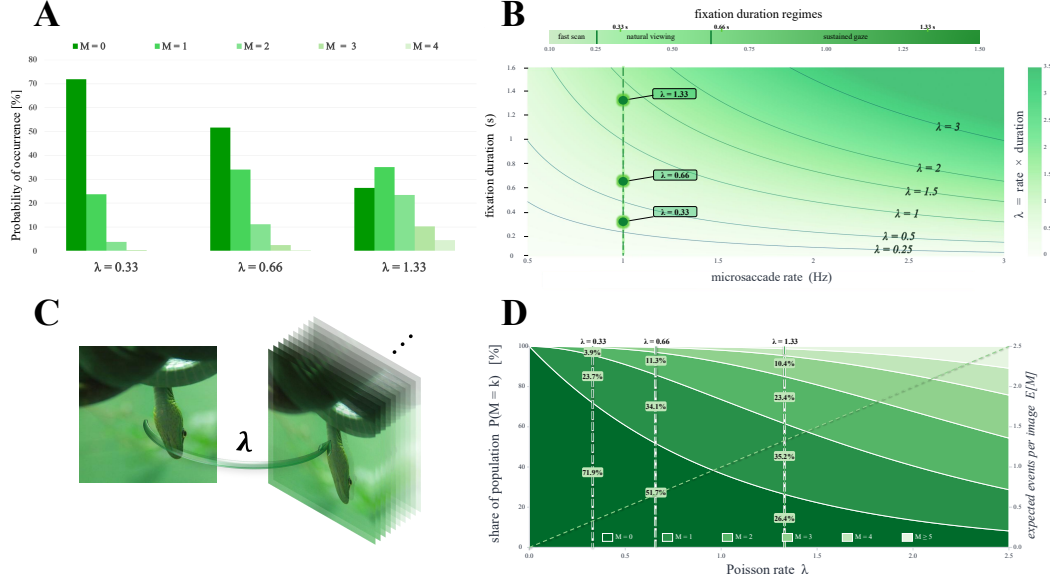


Figure 3: Poisson governance demonstrated. (A) Probability mass function of microsaccade count M at $\lambda \in \{0.33, 0.66, 1.33\}$. (B) Operating points (filled circles) in the (rate, duration) plane at the literature-anchored rate of 1 Hz; color encodes $\lambda = \text{rate} \times \text{duration}$. (C) Augmentation pipeline: $M \sim \text{Poisson}(\lambda)$ is drawn independently per image per epoch. (D) $P(M=k)$ as a continuous function of λ ; vertical markers indicate the three operating points.

$\text{ppd} = N_{\text{px}}/\theta_{\text{FoV}}$, mapping the attested range $[5', 60']$ to $[0.67, 8]$ px on 64×64 and $[2.3, 28]$ px on 224×224 inputs. The upper bound requires no free choice: the 1° microsaccade–saccade boundary corresponds to exactly ppd pixels by construction, ensuring that the largest admissible ghost never exceeds the spatial scale the oculomotor system itself treats as the microsaccade ceiling. Full derivation and cross-dataset validation appear in Appendix B.

3.2 Displacement Statistics: Amplitude Distribution and Direction

Within the $[5', 60']$ range, amplitudes are right-skewed and unimodal on a log scale [Poletti and Rucci, 2016]—the signature of a lognormal, consistent with microsaccades and saccades sharing a common multiplicative-noise substrate [Otero-Millan et al., 2008, Martinez-Conde et al., 2013]. We draw from a truncated lognormal ($\hat{\sigma} = 0.53$, median $19.4'$; fitted and cross-validated in Appendix C) via inverse CDF, so the distribution is exact within the biological bounds with no rejection sampling.

Binocular microsaccades demonstrate a strong cardinal-axis preference: Horizontal events dominate, vertical ones are less frequent and oblique rare [Engbert, 2006, Rolfs, 2009]. We model this with a two-stage sampler: An axis is chosen as horizontal with probability p_h and vertical with $1 - p_h$, followed by equiprobable polarity selection along that axis. A von Mises variate with concentration κ then adds directional jitter around the chosen cardinal direction, greatly reducing, but not eliminating, oblique events. Both $p_h = 0.7193$ and $\kappa = 7.53$ are fitted as described in Section 3 and align with fixation statistics reported in prior studies [Engbert, 2006, Rolfs, 2009, Engbert and Kliegl, 2003]. The bias is defined in the retinal frame [Rolfs, 2009]. Section 4.6 further shows that neither the amplitude distribution nor the directional bias can be replaced by uniform sampling.

3.3 Poisson Occurrence Rate

During steady fixation, microsaccades fire at roughly 1–2 Hz with exponentially distributed inter-event intervals [Engbert and Kliegl, 2003, Rolfs, 2009, Gu et al., 2024]—each event triggering independently once the post-saccadic refractory period clears [Otero-Millan et al., 2008]—so the count $M \sim \text{Poisson}(\lambda=r \cdot T)$ per fixation window is Poisson by construction, with $P(M=k) = \lambda^k e^{-\lambda}/k!$. We fix $r \approx 1$ Hz to the oculomotor consensus [Rolfs, 2009]; T (gaze duration) is the only free

quantity, so λ carries a concrete biological reading rather than arriving as an arbitrary scalar. The empirically optimal λ differs by dataset and is reported in Section 4.4.

3.4 Blend Weight

Unlike a motion blur augmentation that replaces the original with its smeared counterpart, AuGhostmentation superimposes the streak atop the preserved original, with w governing the blend. The biological warrant for which is twofold: intra-saccadic smear is not wholly discarded—cone temporal filtering leaves a transformed signal to propagate downstream [Idrees et al., 2022], and V1 encodes the resulting streak as decodable oriented structure within 50–60 ms of saccade offset [Schweitzer et al., 2025]—and trans-saccadic integration combines the pre-saccadic residual with the incoming foveal signal rather than discarding it [Wolf and Schütz, 2015]. What early visual cortex resolves is therefore a superposition of both states, not a clean snapshot of either. The weighted blend is a direct spatial rendering of this superposition, for which $w = 0.50$ was found most favorable empirically, and its role is verified in Section 4.6.

4 Results

4.1 Clean Accuracy and Corruption Robustness

AuGhostmentation outperforms comparable reference blurs—isotropic Gaussian Blur and directional Motion Blur—as well as the unaugmented baseline on corruption robustness across both datasets, while simultaneously improving clean accuracy on ImageNet-100 and preserving it on Tiny ImageNet (Figure 4, Table 1).

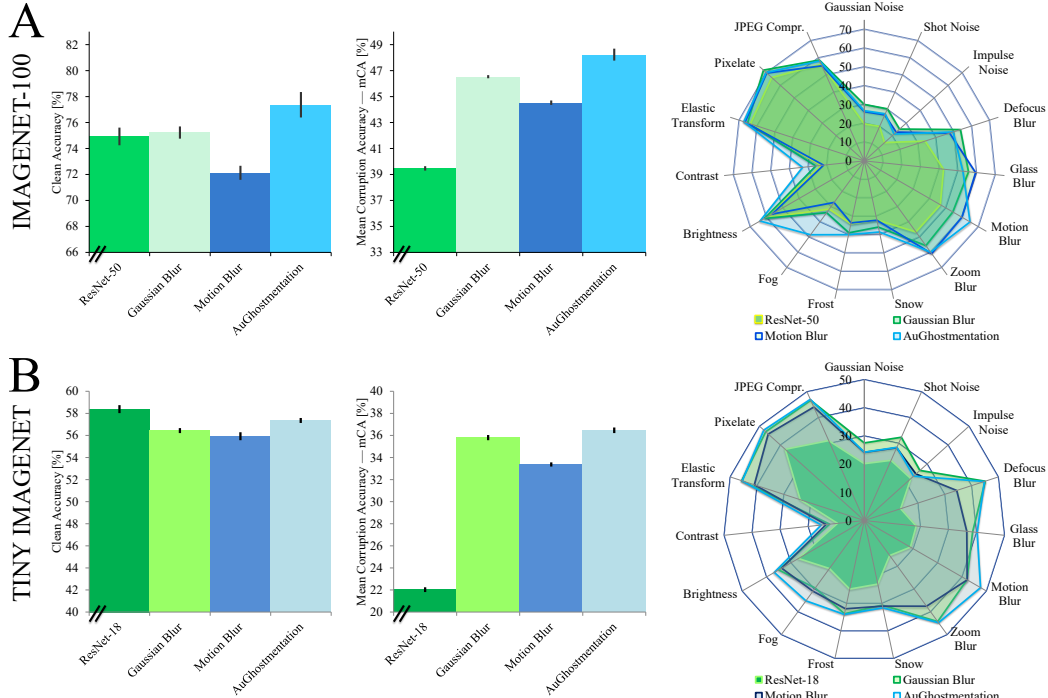


Figure 4: **Clean accuracy and corruption robustness of four training strategies on (A) ImageNet-100 and (B) Tiny ImageNet.** Bar charts (*left*) report Clean Accuracy and mCA [%] for ResNet-50 (A) and ResNet-18 (B) trained without augmentation, with Gaussian Blur, Motion Blur, and AuGhostmentation. Radar charts (*right*) show per-corruption accuracy across 15 corruption types from ImageNet-C/Tiny ImageNet-C, averaged over five severity levels. Error bars denote standard deviation over 3 (A) and 5 (B) independent runs.

Table 1: Clean and corrupted accuracy summary. Values denote mean $\pm$ standard deviation over three seeds. Category columns average all severities and corruptions within each group. Δ mCA is measured relative to the ResNet-50 baseline on ImageNet-100-C.

Method	Clean $\uparrow$	mCA $\uparrow$	Δ mCA	Noise $\uparrow$	Blur $\uparrow$	Weather $\uparrow$	Digital $\uparrow$
ResNet-50	74.9 $\pm$ 0.8	39.5 $\pm$ 0.2	—	18.1 $\pm$ 1.6	42.5 $\pm$ 0.5	39.7 $\pm$ 0.7	52.4 $\pm$ 0.5
Gaussian Blur	75.2 $\pm$ 0.6	46.5 $\pm$ 0.2	+7.1	28.5$\pm$0.6	54.9 $\pm$ 0.8	42.6 $\pm$ 0.4	55.5 $\pm$ 0.3
Motion Blur	72.1 $\pm$ 0.7	44.5 $\pm$ 0.2	+5.1	25.3 $\pm$ 1.1	57.1$\pm$0.7	37.5 $\pm$ 0.6	53.4 $\pm$ 0.7
AuGhostmentation	77.4$\pm$1.2	48.2$\pm$0.6	+8.8	25.1 $\pm$ 0.8	56.9 $\pm$ 2.1	47.9$\pm$1.3	57.2$\pm$0.5

Table 2 extends the comparison to VGG-16 and VOneNet backbones ($n = 3$). Notably, AuGhostmentation compounds its gains further with VOneNet rather than plateauing or collapsing—prepending a retinal prior to a cortical one thus yields additive robustness gains rather than saturation, consistent with the hypothesis that, while the two priors occupy distinct stages of the visual hierarchy, they should nevertheless be compatible together.

Table 2: Clean accuracy and mean Corruption Accuracy (mCA) averages on Tiny ImageNet and ImageNet-100.

Model	Dataset	Clean Acc. (%)	mCA (%)
<i>VGG-16</i>			
Baseline	Tiny ImageNet	46.00 $\pm$ 0.92	17.10 $\pm$ 1.01
+ AuGhostmentation	Tiny ImageNet	43.46 $\pm$ 0.85	27.48 $\pm$ 0.59
<i>VOneNet</i>			
ResNet-18 Baseline	Tiny ImageNet	51.93 $\pm$ 0.28	28.58 $\pm$ 0.54
+ AuGhostmentation	Tiny ImageNet	51.47 $\pm$ 0.49	35.29 $\pm$ 0.48
<i>VOneNet</i>			
ResNet-50 Baseline	ImageNet-100	74.11 $\pm$ 0.37	41.76 $\pm$ 0.55
+ AuGhostmentation	ImageNet-100	76.31 $\pm$ 0.95	46.43 $\pm$ 1.13

4.2 Comparison with Modern Augmentations

Results in Table 3 demonstrate AuGhostmentation against modern augmentation recipes, such as AutoAugment [Cubuk et al., 2019], RandAugment [Cubuk et al., 2020], TrivialAugmentWide [Müller and Hutter, 2021] and AugMix [Hendrycks et al., 2020], under the same training schedules. Prepending AuGhostmentation to each policy raises Tiny ImageNet mCA even further, by 3.5 to 11.0 points, reaching 40.02% with AugMix-JSD (Table 9, Appendix D.5).

Table 3: Augmentation policies alone under the same training recipe (means over seeds): ResNet-18 on Tiny ImageNet ($n = 5$) and ResNet-50 on ImageNet-100 ($n = 3$), accuracy in %.

Policy	Tiny ImageNet		ImageNet-100	
	Clean $\uparrow$	mCA $\uparrow$	Clean $\uparrow$	mCA $\uparrow$
Baseline	58.40	22.02	74.9	39.5
AutoAugment	58.43	30.30	75.04	44.10
AugMix	58.16	32.22	77.04	47.24
RandAugment	59.69	28.06	77.00	45.30
TrivialAugmentWide	58.59	33.83	75.58	47.97
AuGhostmentation	57.41	36.48	77.49	48.20

4.3 Other Architectures

Table 4 reports the same transform, without architecture-specific tuning, on ConvNeXt [Liu et al., 2022], ViT [Dosovitskiy et al., 2021], Swin Transformer [Liu et al., 2021] and MaxViT [Tu et al., 2022], under a standard and a DeiT-style recipe [Touvron et al., 2021] (Appendix D.5).

Table 4: mCA (%) of other backbones without and with AuGhostmentation, under a standard and a DeiT-style recipe (means; Tiny ImageNet $n = 5$, ImageNet-100 $n = 3$). FT: fine-tuned from ImageNet-1K weights.

Dataset	Backbone	Recipe	Baseline	+ AuGhostmentation	Δ mCA
Tiny ImageNet	ConvNeXt	Standard	16.38	26.12	+9.74
	ConvNeXt	DeiT-style	30.67	33.89	+3.22
	ViT (scratch)	DeiT-style	15.15	19.01	+3.86
	ViT (FT)	DeiT-style	53.98	58.31	+4.33
ImageNet-100	Swin-T	Standard	37.08	42.96	+5.88
	Swin-T	DeiT-style	48.50	52.65	+4.15
	MaxViT-T	Standard	49.46	55.06	+5.60
	MaxViT-T	DeiT-style	60.80	63.83	+3.03

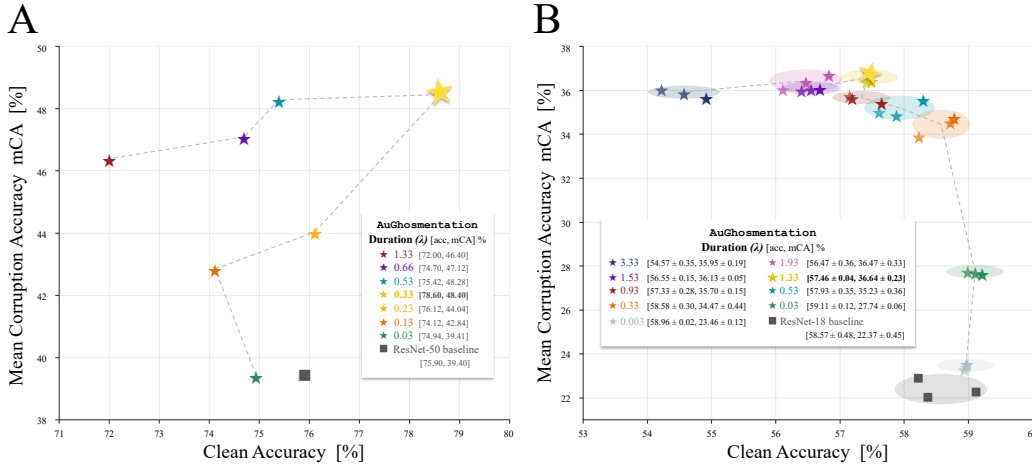


Figure 5: Clean accuracy vs. mean corruption accuracy (mCA) for AuGhostmentation, where the expected number of microsaccades per fixation follows a Poisson distribution with rate λ . (A) ImageNet-100 / ResNet-50: each star corresponds to a single run at a given λ ; the gray square marks the unaugmented baseline. (B) TinyImageNet/ResNet-18: three independent seeds per λ are shown as individual stars; shaded ellipses indicate inter-seed spread. Moderate λ values yield the strongest corruption robustness with minimal clean-accuracy cost.

4.4 Calibration of Fixation Duration

λ is the sole parameter that requires tuning across datasets, a dependence we analogize to the variation of gaze duration in vision, which is likewise contingent on the task at hand [Castelhano et al., 2009]. As demonstrated by the sweep in Figure 5, Tiny ImageNet is best served by a *sustained gaze* duration ($\lambda = 1.33$, $T \approx 1.33$ s), while ImageNet-100 prefers a *fast scan* ($\lambda = 0.33$, $T \approx 0.33$ s). A 64×64 px image is spatially impoverished—each pixel subtends nearly 7.5 arcminutes at 8° FoV—and the visual system, when confronted with such coarse representation, is more likely to dwell. Attesting, we argue, to the biological accounts that posit the fixation durations lengthen on low-acuity or degraded stimuli as the system accumulates evidence it cannot otherwise extract in a single glance [Rayner, 1998, Unema et al., 2005]. Prolonged fixation, thus, means more microsaccades, and AuGhostmentation effectively delivers a richer set of spatially displaced samples of the scene—which is ultimately a form of biological dithering that helps the network integrate what a single low-resolution snapshot could not fully convey. At 224×224 , the image already encodes sufficient high-frequency detail that a brief scan suffices; sustained ghosting at that resolution would begin to smear the very features the network can now afford to exploit—overall performance thus degrades monotonically for $\lambda > 0.33$ on ImageNet-100, as shown in Figure 5A.

Table 5: Effect of horizontal bias weight (p_h) with and without 90° a rotation augmentation. Mean $\pm$ std over 5 seeds.

Rotation	Metric	$p_h = 0.01$	$p_h = 0.29$	$p_h = 0.50$	$p_h = 0.71$	$p_h = 0.99$
None	Clean Acc (%)	53.90 $\pm$ 0.52	53.26 $\pm$ 0.34	53.27 $\pm$ 0.85	53.98 $\pm$ 0.68	54.50 $\pm$ 1.16
	mCA (%)	33.26 $\pm$ 0.68	34.62 $\pm$ 0.42	34.81 $\pm$ 0.41	35.12 $\pm$ 0.44	33.96 $\pm$ 0.74
90°	Clean Acc (%)	49.72 $\pm$ 0.88	49.73 $\pm$ 0.32	48.63 $\pm$ 1.59	48.28 $\pm$ 1.12	49.03 $\pm$ 1.09
	mCA (%)	29.05 $\pm$ 0.78	30.54 $\pm$ 0.25	29.90 $\pm$ 0.72	29.63 $\pm$ 0.43	27.57 $\pm$ 0.39

4.5 Horizontal Bias

Microsaccades are not isotropically distributed; horizontal events dominate, vertical occur at reduced frequency, and oblique are scarce [Engbert and Kliegl, 2003, Rolfs, 2009, Engbert, 2006]. While our findings corroborated the same account—neural networks performed better with horizontal *ghosts*—we wanted to conduct further tests to see if the apparent gain is due to network genuinely exploiting some directional structure or merely benefiting from perturbation energy. We sweep in Table 5 p_h , which denotes horizontal bias, across the full range from near-vertical ($p_h = 0.01$) to near-horizontal ($p_h = 0.99$) and replicate the same sweep with all training and validation images rotated 90°, so that the horizontal axis now becomes vertical and under the hypothesis that a network utilizing a ecological prior would invert its preference accordingly—it *does*.

4.6 Ablation Tests

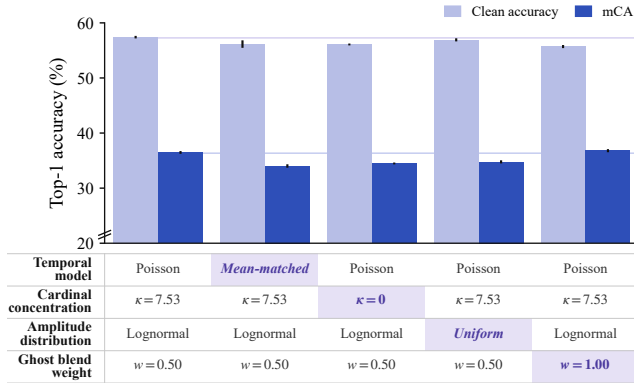


Figure 6: **Single-parameter ablations of AuGhostmentation’s biological priors on Tiny ImageNet (ResNet-18).** Clean accuracy (*light*) and mCA (*dark*); each column changes only the highlighted component, where lines mark the full configuration. *Mean-matched*: $M = 1$ or 2 with probabilities 0.67 and 0.33, matching the Poisson mean $\mathbb{E}[M] = 1.33$; *Uniform*: amplitudes drawn uniformly over the same range; $\kappa = 0$: uniform directions; $w = 1.00$: original frame dropped from the blend.

We ablate each biological prior in turn on Tiny ImageNet, holding all other components at their fitted defaults (Figure 6). The lognormal amplitude is replaced with a uniform draw over the same biological range to isolate the contribution of distributional shape from that of support. The cardinal directional prior is collapsed to $\kappa = 0$ to test whether the axis-concentration is meaningful. The Poisson model is replaced with a mean-matched Bernoulli draw. The blend weight is set as a test of the *masking* theory: Setting $w = 1.0$ strips the pre-saccadic original from the training sample—the very frame that, in the brain, masks and suppresses the smear [Duyck et al., 2016, Balsdon et al., 2018]. Corruption accuracy is unaffected, yet

clean accuracy falls, indicating robustness as being carried by the directional kernel and clean accuracy by the blend.

4.7 Assessment of Behavioral Bias: Stylized ImageNet

Following prior stylization and cue-conflict work [Geirhos et al., 2019, Huang and Belongie, 2017], we report stylized top-1 and matched-clean accuracy to probe *why* AuGhostmentation outperforms competing models. AdaIN transfer reveals a clear texture-to-shape shift: AuGhostmentation achieves higher shape bias than plain ResNet-50 or combined with Gaussian Blur versus Motion Blur. Notably, while Gaussian Blur achieves comparable corruption robustness as AuGhostmentation

Table 6: Cue-conflict shape-bias index (SB) and forced-choice coverage on ImageNet-100 [Geirhos et al., 2019] ($n = 3$). See Appendix E for protocol details.

Model	Shape Bias (SB)	Coverage
AuGhostmentation	78.9$\pm$0.1%	27.1$\pm$0.7%
Gaussian Blur	75.9 $\pm$ 0.9%	24.6 $\pm$ 0.1%
ResNet-50 (baseline)	70.7 $\pm$ 0.8%	23.0 $\pm$ 0.1%
Motion Blur	70.3 $\pm$ 3.1%	21.3 $\pm$ 2.1%

Table 7: WikiArt/AdaIN stylized top-1 and matched top-1 accuracy across style-strength levels α . Mean $\pm$ std over ($n = 3$). See Appendix E for protocol details.

Model	Top-1 (%)				Matched Top-1 (%)			
	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1.00$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1.00$
AuGhostmentation	68.3$\pm$0.9	56.6$\pm$1.2	37.0$\pm$1.2	18.5$\pm$0.8	83.7$\pm$0.1	68.9$\pm$0.6	44.7$\pm$0.8	22.0$\pm$0.7
Gaussian Blur	65.9 $\pm$ 0.7	53.2 $\pm$ 0.2	33.3 $\pm$ 0.6	16.4 $\pm$ 0.9	81.9 $\pm$ 0.9	65.7 $\pm$ 1.6	40.9 $\pm$ 1.6	19.6 $\pm$ 1.5
ResNet-50 (baseline)	65.1 $\pm$ 0.2	51.2 $\pm$ 0.2	30.5 $\pm$ 0.7	14.1 $\pm$ 0.4	81.8 $\pm$ 0.4	64.1 $\pm$ 0.6	37.7 $\pm$ 0.9	17.1 $\pm$ 0.5
Motion Blur	62.4 $\pm$ 3.3	48.9 $\pm$ 3.5	29.0 $\pm$ 2.9	13.7 $\pm$ 1.8	79.4 $\pm$ 2.2	61.8 $\pm$ 2.8	36.2 $\pm$ 2.8	16.7 $\pm$ 1.8

(Figure 4), it remains far less shape-biased, suggestive of a different robustness mechanism at play. The contrast with Motion Blur is likewise instructive: Both are loosely, "a type of motion blur," yet Motion Blur demonstrably fails to increase neither stylized accuracy or shape bias, while AuGhostmentation improves both. A parsimonious explanation is that standard motion blur applies a spatially uniform linear convolution, leaving the relative salience of texture and shape cues unchanged; AuGhostmentation, instead, via simulated fixation jitter with biologically realistic saccade characteristics, selectively disrupts fine-grained texture along an axis while preserving coarser shape structure—a fundamentally different inductive bias.

5 Conclusion

We introduced AuGhostmentation, a biologically grounded training augmentation that simulates the intra-saccadic retinal smear from empirically fitted microsaccade statistics, and showed that it improves corruption robustness and clean accuracy simultaneously—a combination most augmentations trade against each other. Some of our findings proved more instructive than anticipated: The blend weight carries the clean-accuracy gain by reinstating the masking frame the brain uses to suppress the smear, putatively attesting to the biological trans-saccade accounts; the directional prior inverts under 90° rotation, confirming a genuine spatial bias rather than a perturbation-energy effect; and AuGhostmentation compounds additively with VOneNet, suggesting retinal and cortical priors occupy distinct enough stages to act in concert. The fact that merely four empirically calibrated parameters, acting in tandem, suffice to produce effects this robust may be the most compelling evidence that the oculomotor system has converged on a spatial representation of genuine and lasting utility. Implementations worth further pursuing include luminance attenuation during saccades, drift-based fixational corrections and predictive coding of pre- and post-saccadic frames—well-characterized oculomotor phenomena the present architecture leaves unmodeled—and tests of directional or other apparent biases on varying architectures where improvements follow.

Acknowledgments and Disclosure of Funding

This study was partially supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under Grant No. 125E068 and Boğaziçi University Scientific Research Projects (BAP) under Grant No. 25A01SUP1.

References

David H. Hubel and Torsten N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962. doi:

- 10.1113/jphysiol.1962.sp006837.
- David H. Hubel and Torsten N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243, 1968. doi: 10.1113/jphysiol.1968.sp008455.
- Kunihiko Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20(3–4):121–136, 1975. doi: 10.1007/BF00342633.
- Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980. doi: 10.1007/BF00344251.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. doi: 10.1038/nature14539. URL <https://doi.org/10.1038/nature14539>.
- Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=HJz6tiCqYm>.
- Robert Geirhos, Carlos R. M. Temme, Jonas Rauber, Heiko H. Schütt, Matthias Bethge, and Felix A. Wichmann. Generalisation in humans and deep neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- Fabian H. Sinz, Xaq Pitkow, Jacob Reimer, Matthias Bethge, and Andreas S. Tolias. Engineering a less artificial intelligence. *Neuron*, 103(6):967–979, 2019. doi: 10.1016/j.neuron.2019.08.034.
- Charles F. Cadieu, Ha Hong, Daniel L. K. Yamins, Nicolas Pinto, Diego Ardila, Ethan A. Solomon, Najib J. Majaj, and James J. DiCarlo. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLOS Computational Biology*, 10(12):e1003963, 2014. doi: 10.1371/journal.pcbi.1003963.
- Daniel L. K. Yamins and James J. DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356–365, 2016. doi: 10.1038/nn.4244. URL <https://www.nature.com/articles/nn.4244>.
- Radoslaw Martin Cichy, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, 6:27755, 2016. doi: 10.1038/srep27755.
- Blake A. Richards, Timothy P. Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, Colleen J. Gillon, Danijar Hafner, Adam Kepecs, Nikolaus Kriegeskorte, Peter Latham, Grace W. Lindsay, Kenneth D. Miller, Richard Naud, Christopher C. Pack, Panayiota Poirazi, Pieter R. Roelfsema, João Sacramento, Andrew Saxe, Benjamin Scellier, Anna C. Schapiro, Walter Senn, Greg Wayne, Daniel Yamins, Friedemann Zenke, Joel Zylberberg, Denis Thérien, and Konrad P. Kording. A deep learning framework for neuroscience. *Nature Neuroscience*, 22(11):1761–1770, 2019. doi: 10.1038/s41593-019-0520-2.
- Joel Dapello, Tiago Marques, Martin Schrimpf, Franziska Geiger, David D. Cox, and James J. DiCarlo. Simulating a primary visual cortex at the front of cnns improves robustness to image perturbations. In *Advances in Neural Information Processing Systems*, volume 33, pages 13073–13087, 2020.
- Suayb S. Arslan. Artificial human intelligence: The role of humans in the development of next generation AI. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 10(1):4–20, 2026. doi: 10.1109/TETCI.2025.3645582.
- Avinash Baidya, Joel Dapello, James J DiCarlo, and Tiago Marques. Combining different V1 brain model variants to improve robustness to image corruptions in CNNs. *arXiv preprint arXiv:2110.10645*, 2021.
- Benjamin D. Evans, Gaurav Malhotra, and Jeffrey S. Bowers. Biological convolutions improve DNN robustness to noise and generalisation. *Neural Networks*, 148:96–110, 2022. doi: 10.1016/j.neunet.2021.12.005.

- Lucas Piper, Arlindo L. Oliveira, and Tiago Marques. Explicitly modeling subcortical vision with a neuro-inspired front-end improves CNN robustness. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL <https://arxiv.org/abs/2506.03089>.
- Lucas Piper, Arlindo L. Oliveira, and Tiago Marques. Explicitly modeling pre-cortical vision with a neuro-inspired front-end improves CNN robustness, 2024. URL <https://arxiv.org/abs/2409.16838>.
- Harshitha Machiraju, Oh-Hyeon Choung, Pascal Frossard, and Michael H. Herzog. Bio-inspired robustness: A review. *arXiv preprint arXiv:2103.09265*, 2021.
- Eileen Kowler. Eye movements: the past 25 years. *Vision Research*, 51(13):1457–1483, 2011. doi: 10.1016/j.visres.2010.12.014.
- Kari L. Hoffman, Michelle C. Dragan, Timothy K. Leonard, Cristiano Micheli, Rodrigo Montefusco-Siegmund, and Taufik A. Valiante. Saccades during visual exploration align hippocampal 3–8 Hz rhythms in human and non-human primates. *Frontiers in Systems Neuroscience*, 7:43, 2013. doi: 10.3389/fnsys.2013.00043.
- Zoe Stearns and Martina Poletti. Temporal dynamics of peri-microsaccadic modulations within the foveola. *eLife*, February 2026. doi: 10.7554/elife.109144.1. URL <http://dx.doi.org/10.7554/eLife.109144.1>. Reviewed Preprint, version 1.
- Martin Rolfs. Microsaccades: Small steps on a long way. *Vision Research*, 49(20):2415–2441, 2009. ISSN 0042-6989. doi: 10.1016/j.visres.2009.08.010. URL <https://www.sciencedirect.com/science/article/pii/S0042698909003691>.
- Ralf Engbert and Reinhold Kliegl. Microsaccades uncover the orientation of covert attention. *Vision Research*, 43(9):1035–1045, 2003. ISSN 0042-6989. doi: 10.1016/S0042-6989(03)00084-1. URL <https://www.sciencedirect.com/science/article/pii/S0042698903000841>.
- Susana Martinez-Conde, Stephen L. Macknik, and David H. Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3):229–240, 2004. doi: 10.1038/nrn1348. URL <https://doi.org/10.1038/nrn1348>.
- Susana Martinez-Conde, Jorge Otero-Millan, and Stephen L. Macknik. The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14(2): 83–96, 2013. doi: 10.1038/nrn3405.
- David Melcher and Carol L. Colby. Trans-saccadic perception. *Trends in Cognitive Sciences*, 12(12): 466–473, 2008. doi: 10.1016/j.tics.2008.09.003.
- Marc A. Sommer and Robert H. Wurtz. Visual perception and corollary discharge. *Perception*, 37(3): 408–418, 2008. doi: 10.1068/p5873.
- Robert H. Wurtz, Wilsaan M. Joiner, and Rebecca A. Berman. Neuronal mechanisms for visual stability: progress and problems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1564):492–503, 2011. doi: 10.1098/rstb.2010.0186.
- Harold E. Bedell, Susana T. L. Chung, and Saumil S. Patel. Attenuation of perceived motion smear during vergence and pursuit tracking. *Vision Research*, 44(9):895–902, April 2004. doi: 10.1016/j.visres.2003.11.006.
- Marianne Duyck, Thérèse Collins, and Mark Wexler. Masking the saccadic smear. *Journal of Vision*, 16(10):1, 08 2016. ISSN 1534-7362. doi: 10.1167/16.10.1. URL <https://doi.org/10.1167/16.10.1>.
- Frédéric Crevecoeur and Konrad P Kording. Saccadic suppression as a perceptual consequence of efficient sensorimotor estimation. *eLife*, 6:e25073, may 2017. ISSN 2050-084X. doi: 10.7554/eLife.25073. URL <https://doi.org/10.7554/eLife.25073>.

- T. Balsdon, R. Schweitzer, T.L. Watson, and M. Rolfs. All is not lost: Post-saccadic contributions to the perceptual omission of intra-saccadic streaks. *Consciousness and Cognition*, 64:19–31, 2018. ISSN 1053-8100. doi: 10.1016/j.concog.2018.05.004. URL <https://www.sciencedirect.com/science/article/pii/S1053810018300473>. Visual Experience and Guidance of Action: A Tribute to Bruce Bridgeman.
- Richard Schweitzer and Martin Rolfs. Intrасaccadic motion streaks jump-start gaze correction. *Science Advances*, 7(30):eabf2218, 2021. doi: 10.1126/sciadv.abf2218. URL <https://www.science.org/doi/abs/10.1126/sciadv.abf2218>.
- Richard Sylvester, John-Dylan Haynes, and Geraint Rees. Saccades differentially modulate human LGN and V1 responses in the presence and absence of visual stimulation. *Current Biology*, 15(1): 37–41, 2005. doi: 10.1016/j.cub.2004.12.061.
- Richard Sylvester and Geraint Rees. Extraretinal saccadic signals in human LGN and early retinotopic cortex. *NeuroImage*, 30(1):214–219, 2006. doi: 10.1016/j.neuroimage.2005.09.014.
- Ethel Martin. Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81(12): 899–917, 1974. doi: 10.1037/h0037368.
- Frances C. Volkman. Human visual suppression. *Vision Research*, 26(9):1401–1416, 1986. doi: 10.1016/0042-6989(86)90164-1.
- Mark R. Diamond, John Ross, and M. Concetta Morrone. Extraretinal control of saccadic suppression. *Journal of Neuroscience*, 20(9):3449–3455, 2000. doi: 10.1523/JNEUROSCI.20-09-03449.2000.
- John Ross, M. Concetta Morrone, Michael E. Goldberg, and David C. Burr. Changes in visual perception at the time of saccades. *Trends in Neurosciences*, 24(2):113–121, 2001. doi: 10.1016/S0166-2236(00)01685-4.
- Paola Binda and Maria Concetta Morrone. Vision during saccadic eye movements. *Annual Review of Vision Science*, 4:193–213, 2018. doi: 10.1146/annurev-vision-091517-034317.
- Fergus W. Campbell and Robert H. Wurtz. Saccadic omission: Why we do not see a grey-out during a saccadic eye movement. *Vision Research*, 18(10):1297–1303, 1978. doi: 10.1016/0042-6989(78)90219-5.
- Eric Castet, Sébastien Jeanjean, and Guillaume S. Masson. Motion perception of saccade-induced retinal translation. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15159–15163, 2002. doi: 10.1073/pnas.232377199.
- Jorge Otero-Millan, Xoana G. Troncoso, Stephen L. Macknik, Ignacio Serrano-Pedraza, and Susana Martinez-Conde. Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of Vision*, 8(14):21–21, 2008. doi: 10.1167/8.14.21.
- Ziad M. Hafed, Chih-Yang Chen, and Xiaoguang Tian. Vision, perception, and attention through the lens of microsaccades: Mechanisms and implications. *Frontiers in Systems Neuroscience*, 9:167, 2015. doi: 10.3389/fnsys.2015.00167.
- Barbara Landau, Linda B. Smith, and Susan S. Jones. The importance of shape in early lexical learning. *Cognitive Development*, 3(3):299–321, 1988. doi: 10.1016/0885-2014(88)90014-7.
- Irving Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987. doi: 10.1037/0033-295X.94.2.115.
- Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=Bygh9j09KX>.
- Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422, 1998. doi: 10.1037/0033-2909.124.3.372.

- Tom Foulsham, Alan Kingstone, and Geoffrey Underwood. Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, 48(17):1777–1790, 2008. doi: 10.1016/j.visres.2008.05.018.
- Benjamin W. Tatler and Benjamin T. Vincent. Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2):1–18, 2008. doi: 10.16910/jemr.2.2.5.
- A. Terry Bahill, Michael R. Clark, and Lawrence Stark. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24(3–4):191–204, 1975. doi: 10.1016/0025-5564(75)90075-9.
- Henry Griffith, Dillon Lohr, Evgeny Abdulin, and Oleg V. Komogortsev. GazeBase, a large-scale, multi-stimulus, longitudinal eye movement dataset. *Scientific Data*, 8(1):184, 2021. doi: 10.1038/s41597-021-00959-y. URL <https://doi.org/10.1038/s41597-021-00959-y>.
- B. L. Zuber, L. Stark, and G. Cook. Microsaccades and the velocity–amplitude relationship for saccadic eye movements. *Science*, 150(3702):1459–1460, 1965. doi: 10.1126/science.150.3702.1459.
- Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo. Brain-Score: Which artificial neural network for object recognition is most brain-like? Technical report, bioRxiv, 2018.
- Tiago Marques, Martin Schrimpf, and James J. DiCarlo. Hierarchical neural network models that more closely match primary visual cortex tend to better explain higher level visual cortical responses. Cosyne Abstracts, 2020. Poster I-12, Denver, CO, USA.
- Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, Matthias Bethge, and Alexander S Ecker. Deep convolutional models improve predictions of macaque v1 responses to natural images. *PLOS Computational Biology*, 15(4):e1006897, 2019. doi: 10.1371/journal.pcbi.1006897.
- Martina Poletti and Michele Rucci. A compact field guide to the study of microsaccades: Challenges and functions. *Vision Research*, 118:83–97, 2016. doi: 10.1016/j.visres.2015.01.018.
- Ralf Engbert. Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. In *Progress in Brain Research*, volume 154, pages 177–192. Elsevier, 2006. doi: 10.1016/S0079-6123(06)54009-9.
- Quan Gu, Qikai Zhang, Yueming Han, Pengfei Li, Zaifeng Gao, and Mowei Shen. Microsaccades reflect attention shifts: a mini review of 20 years of microsaccade research. *Frontiers in Psychology*, 15:1364939, 2024. ISSN 1664-1078. doi: 10.3389/fpsyg.2024.1364939. URL <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1364939>.
- Saad Idrees, Matthias-Philipp Baumann, Maria M. Korympidou, Timm Schubert, Alexandra Kling, Katrin Franke, Ziad M. Hafed, Felix Franke, and Thomas A. Münch. Suppression without inhibition: how retinal computation contributes to saccadic suppression. *Communications Biology*, 5(1):692, 2022. doi: 10.1038/s42003-022-03526-2. URL <https://doi.org/10.1038/s42003-022-03526-2>.
- Richard Schweitzer, Thomas Seel, Jörg Raisch, and Martin Rolfs. Early visual signatures and benefits of intra-saccadic motion streaks. *PLOS Computational Biology*, 21(9):e1013544, 09 2025. doi: 10.1371/journal.pcbi.1013544. URL <https://doi.org/10.1371/journal.pcbi.1013544>.
- Christian Wolf and Alexander C. Schütz. Trans-saccadic integration of peripheral and foveal feature information is close to optimal. *Journal of Vision*, 15(16):1, 2015. doi: 10.1167/15.16.1.
- Ekin D. Cubuk, Barret Zoph, Dandelion Mané, Vijay Vasudevan, and Quoc V. Le. AutoAugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 113–123, 2019. doi: 10.1109/CVPR.2019.00020.

- Ekin Dogus Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. RandAugment: Practical automated data augmentation with a reduced search space. In *Advances in Neural Information Processing Systems*, volume 33, pages 18613–18624, 2020.
- Samuel G. Müller and Frank Hutter. TrivialAugment: Tuning-free yet state-of-the-art data augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 754–762, 2021. doi: 10.1109/ICCV48922.2021.00081.
- Dan Hendrycks, Norman Mu, Ekin D. Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. AugMix: A simple data processing method to improve robustness and uncertainty. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=S1gmrxFvB>.
- Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11966–11976, 2022. doi: 10.1109/CVPR52688.2022.01167.
- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=YicbFdNTTy>.
- Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, 2021. doi: 10.1109/ICCV48922.2021.00986.
- Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. MaxViT: Multi-axis vision transformer. In *Computer Vision – ECCV 2022*, volume 13684 of *Lecture Notes in Computer Science*, pages 459–479, Cham, 2022. Springer. doi: 10.1007/978-3-031-20053-3_27.
- Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 10347–10357, 2021.
- Monica S. Castelhana, Michael L. Mack, and John M. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):6, 2009. doi: 10.1167/9.3.6.
- Pieter J. A. Unema, Sebastian Pannasch, Markus Joos, and Boris M. Velichkovsky. Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3):473–494, 2005. doi: 10.1080/13506280444000409.
- Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1510–1519, Oct 2017. doi: 10.1109/ICCV.2017.167. URL <https://arxiv.org/abs/1703.06868>.
- Shivany Y. Condor Montes, Daniel Bennett, Ethan Bensinger, Lakshmisahithi Rani, Younes Sherkat, Chao Zhao, Zachary Helft, Austin Roorda, Ari J. Green, and Christy K. Sheehy. Characterizing fixational eye motion variance over time as recorded by the tracking scanning laser ophthalmoscope. *Translational Vision Science & Technology*, 11(2):35, 2022. doi: 10.1167/tvst.11.2.35.
- Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2):125, 2020. doi: 10.3390/info11020125.
- Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2018. URL <https://openreview.net/forum?id=r1Ddp1-Rb>.

- Sangdo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. CutMix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6022–6031, 2019. doi: 10.1109/ICCV.2019.00612.
- Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13001–13008, 2020. doi: 10.1609/aaai.v34i07.7000.
- Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5987–5995, 2017. doi: 10.1109/CVPR.2017.634.
- Dan Hendrycks, Andy Zou, Mantas Mazeika, Leonard Tang, Bo Li, Dawn Song, and Jacob Steinhardt. PixMix: Dreamlike pictures comprehensively improve safety measures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16762–16771, 2022. doi: 10.1109/CVPR52688.2022.01628.
- Naoto Inoue. pytorch-AdaIN: PyTorch implementation of arbitrary style transfer in real-time with adaptive instance normalization. <https://github.com/naoto0804/pytorch-AdaIN>, 2019.
- Gaëlle Nicolas, Eric Castet, Adrien Rabier, Emmanuelle Kristensen, Michel Dojat, and Anne Guérin-Dugué. Neural correlates of intra-saccadic motion perception. *Journal of Vision*, 21(11):19, 2021. doi: 10.1167/jov.21.11.19.

Appendix

A A Mathematical Formulation of AuGhostmentation

For the k -th microsaccadic event, let the sampled amplitude and direction be a_k and $\mathbf{u}_k$, respectively. The resulting displacement vector would be given by $\mathbf{d}_k = a_k \mathbf{u}_k$. We define the kernel radius and set the number of rasterization samples as follows,

$$R_k = \max(2, \text{round}(a_k)), \quad M_k = 2R_k,$$

A discrete motion kernel $H_k \in \mathbb{R}^{(2R_k+1) \times (2R_k+1)}$ is then constructed by uniformly sampling points along the line segment representing the microsaccadic trajectory, i.e., from the kernel center (R_k, R_k) to the displaced endpoint $(R_k + d_{x,k}, R_k + d_{y,k})$. More specifically, for $t_m = \frac{m}{M_k-1}$, $m = 0, 1, \dots, M_k - 1$, the intermediate count map is defined as

$$\tilde{H}_k(i, j) = \sum_{m=0}^{M_k-1} \mathbf{1}[(i, j) = (\text{round}(R_k + t_m d_{y,k}), \text{round}(R_k + t_m d_{x,k}))]$$

Intuitively, $\tilde{H}_k(i, j)$ records how many sampled trajectory points fall onto each discrete pixel location. The final normalized kernel is obtained as

$$H_k(i, j) = \frac{\tilde{H}_k(i, j)}{\sum_{i', j'} \tilde{H}_k(i', j')},$$

Finally, the sequential augmentation step is applied as

$$I_k = (1 - w)I_{k-1} + w(I_{k-1} * H_k).$$

where w controls the blending strength of the microsaccadic motion effect.

B Field of View (FoV)

θ_{FoV} is the single design choice that bridges biological angular units and CNN pixel space. Given an image of N_{px} pixels spanning θ_{FoV} degrees, the pixels-per-degree conversion is $\text{ppd} = N_{\text{px}}/\theta_{\text{FoV}}$, and the biologically attested amplitude range $[5', 60']$ maps to $[5/60, 1] \cdot \text{ppd}$ pixels. The upper bound is exact: the 1° microsaccade ceiling corresponds to exactly ppd pixels by construction, so no biological constraint is relaxed. We adopt $\theta_{\text{FoV}} = 8^\circ$ following Dapello et al. [2020], ensuring that the spatial scale of AuGhostmentation is commensurate with the downstream VOneNet stage that was calibrated under the same assumption.

Figure 7 illustrates the consequence of this choice across the full resolution range from 32 to 512 pixels. Panel A shows the truncated lognormal amplitude density in pixel space: as resolution increases the distribution shifts rightward, with the median displacement growing from 1.3 px at 32 px to 20.7 px at 512 px, while the right-skewed shape is preserved throughout. Panel B replots the same densities on a logarithmic amplitude axis, where a lognormal distribution manifests as a symmetric bell curve. The near-perfect symmetry across all resolutions confirms that the lognormal parameterisation ($\hat{\sigma} = 0.53$, median $19.4'$) is a faithful generative model of the pixel-space displacements regardless of input size, and that the FoV assumption introduces no distributional distortion.

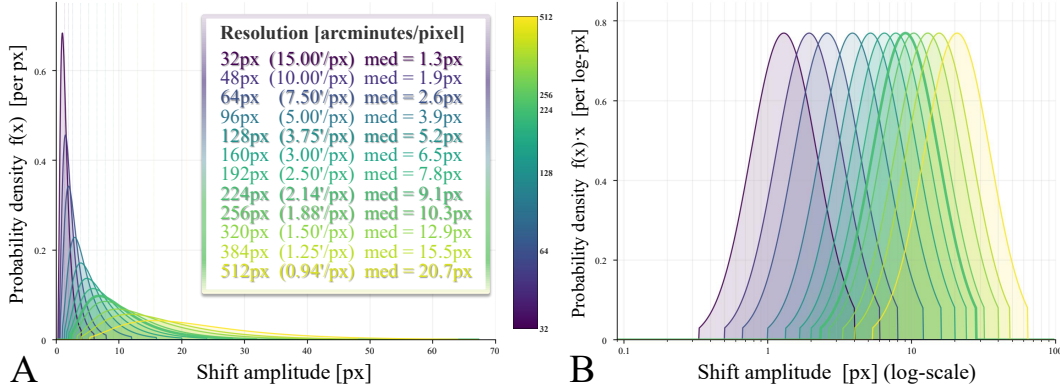


Figure 7: **Truncated lognormal amplitude distributions in pixel space across image resolutions** ($\theta_{\text{FoV}} = 8^\circ$). **(A)** Probability density $f(x)$ in pixels for resolutions from 32 to 512 px (cool to warm colormap); the legend reports the arcminutes-per-pixel scale and the resulting median pixel displacement for each resolution. **(B)** The same distributions on a log-scale amplitude axis, where a lognormal is expected to appear as a symmetric bell. The near-perfect symmetry across all resolutions confirms that the fitted lognormal ($\hat{\sigma} = 0.53$, median $19.4'$) faithfully parameterises the pixel-space displacement distribution independent of input size.

C Oculomotor Parameter Derivation from GazeBase v2.0

C.1 Dataset

We used **GazeBase v2.0** [Griffith et al., 2021] (CC-BY-4.0): 12,334 monocular recordings from 322 college-aged participants across 9 longitudinal rounds, collected with an SR Research EyeLink 1000 at $f_s = 1,000$ Hz, 550 mm viewing distance, chin-and-forehead rest, nine-point calibration (mean error $< 0.5^\circ$). Of the seven available tasks, we used *exclusively* the **fixation task** (FXS): 15 seconds of steady gaze on a stationary bull’s-eye—the only condition free of pursuit, reading, and task-driven saccades. High-precision alternatives such as Condor Montes et al. [2022] (14 subjects, TSLO) and Poletti and Rucci [2016] (~ 10 subjects, DPI) offer superior spatial resolution but are far too small for stable population-level fitting of p_h and κ , and are not publicly available.

C.2 Microsaccade Detection

We applied the Engbert–Kliegl velocity-threshold algorithm [Engbert and Kliegl, 2003]. Instantaneous velocity is estimated by the five-point central difference:

$$\dot{x}_i = \frac{(x_{i+1} + x_{i+2}) - (x_{i-1} + x_{i-2})}{6} \cdot f_s, \quad (1)$$

and identically for $\dot{y}_i$. A per-segment adaptive threshold is derived from the median-based noise floor:

$$\hat{\sigma}_x = \sqrt{\text{median}(\dot{x}^2) - [\text{median}(\dot{x})]^2}, \quad \eta_x = \lambda \hat{\sigma}_x, \quad \lambda = 6, \quad (2)$$

with the elliptical supra-threshold criterion $(\dot{x}_i/\eta_x)^2 + (\dot{y}_i/\eta_y)^2 > 1$. Contiguous supra-threshold runs are retained as microsaccades if: duration ≥ 6 ms, amplitude $5' \leq A \leq 60'$ [Zuber et al., 1965, Martinez-Conde et al., 2013], and inter-event gap ≥ 50 ms [Otero-Millan et al., 2008]. The pipeline yielded $N = 28,916$ events across 322 subjects and 1,755 recordings.

C.3 Directional Parameters ($\hat{p}_h, \hat{\kappa}$)

Direction $\theta_i = \text{atan2}(\Delta y_i, \Delta x_i)$. Events within $\pm 45^\circ$ of either horizontal axis are classified horizontal, giving $\hat{p}_h = N_H/N = 0.7193$. A four-component von Mises mixture with centres $\mu = (0, \frac{\pi}{2}, \pi, \frac{3\pi}{2})$, fixed weights from $\hat{p}_h$, and shared concentration κ is fitted by maximising

$$\ell(\kappa) = \sum_{i=1}^N \log \left[\sum_{k=1}^4 w_k \cdot f_{\text{vM}}(\theta_i \mid \mu_k, \kappa) \right] \quad (3)$$

via grid search followed by Brent-method refinement, yielding $\hat{\kappa} = 7.53$, 95% bootstrap CI [7.38, 7.68] ($B=200$).

C.4 Amplitude Parameters ($\hat{\sigma}$, median) and Cross-Validation

A truncated lognormal $\mathcal{LN}(\mu_\ell, \sigma^2)$ on $[5', 60']$ is fit to the empirical amplitude histogram by MLE, yielding $\hat{\sigma} = 0.53$ and median $e^{\mu_\ell} = 19.4'$ (Table 8). The $\sim 2'$ excess over the TSLO median of Condor Montes et al. [2022] (17.4') is expected: the EyeLink 1000 positional noise (~ 15 – $30'$) inflates measured displacements near the detection floor. Both noise-induced biases are conservative—EyeLink *dilutes* κ toward isotropy and *inflates* the amplitude median—so all four fitted values understate rather than overstate the biological signal.

Source	Statistic	Value (')
GazeBase v2.0 (EyeLink)	Median	19.4
Condor Montes et al. [2022] (TSLO)	Median	17.4
Poletti and Rucci [2016] (DPI)	63rd pct.	20
Poletti and Rucci [2016] (DPI)	90th pct.	31
Ours ($\hat{\sigma} = 0.53$, median= 19.4')	$F(20')$	0.58

Table 8: Amplitude cross-validation across three independent measurement modalities. $F(20')$ denotes the fraction of microsaccades our fitted lognormal predicts below $20'$; the DPI data of Poletti and Rucci [2016] places the 63rd percentile at $20'$, confirming agreement to within ~ 5 percentage points.

D Training Setup

D.1 Tiny ImageNet (ResNet-18)

We adopt the training recipe and architecture modifications of Baidya et al. [2021]. The backbone is a ResNet-18 variant in which the stride of the first convolutional layer is reduced from two to one while the maxpool stride is retained at two, yielding a combined stride of two commensurate with the VOneBlock; this modification raises clean accuracy from 50.45% (standard Torchvision ResNet-18) to 58.93% on Tiny ImageNet. Tiny ImageNet contains 100,000 images across 200 classes (64×64 , 500 training and 50 validation images per class). Models are trained for 60 epochs; all remaining hyperparameters follow Baidya et al. [2021] without modification. AuGhostmentation is prepended to the standard augmentation pipeline as a probabilistic frontend transform. All Tiny ImageNet runs use multiple seeds and report mean $\pm$ standard deviation.

D.2 Tiny ImageNet (VGG-16)

To provide a comparable baseline against the VOneNet family on Tiny ImageNet, we construct a *fair* VGG-16 variant whose spatial footprint matches that of the VOneBlock-modified ResNet-18. The first convolutional layer is replaced with a 7×7 , stride-2, padding-3 layer (bias enabled, no BatchNorm), reducing the combined spatial stride to two in a manner commensurate with the VOneBlock. The final MaxPool in the feature extractor is removed, and the global average pool is replaced with an identity, yielding a 2×2 spatial map ($512 \times 2 \times 2 = 2048$ features). The classifier head is re-specified as three fully connected layers: $2048 \rightarrow 4096 \rightarrow 2048 \rightarrow 200$ with ReLU activations and dropout between each pair.

Training uses SGD with learning rate 0.01, momentum 0.9, and weight decay 5×10^{-4} for 100 epochs with a batch size of 128. The learning rate is governed by ReduceLROnPlateau (factor 0.1, patience 5, monitoring validation loss). The training pipeline applies a ReflectAffine transform (rotation $\pm 30^\circ$, translation $\pm 5\%$, scale 1.0–1.2, bilinear interpolation with reflection padding), random horizontal flip, AuGhostmentation ($\theta_{\text{FoV}} = 8^\circ$, $p_h = 0.7193$, $\kappa = 7.5347$), a centre crop to 64×64 , and normalisation by $[0.5, 0.5, 0.5]$. Validation images are normalized only.

D.3 ImageNet-100 (ResNet-50)

We adopt the training procedure of Dapello et al. [2020] without modification, applied to a 100-class subset of ImageNet. The backbone is a standard, unmodified ResNet-50. Training images are preprocessed with random crop to 224×224 and random horizontal flip; validation images use a central crop. Pixel values are normalised by subtracting and dividing by $[0.5, 0.5, 0.5]$. Optimisation uses SGD with momentum 0.9, weight decay 10^{-4} , and a step learning rate schedule (initial rate 0.1, divided by 10 every 20 epochs) for 70 epochs with a batch size of 256. AuGhostmentation is applied as a frontend transform under the same pipeline.

D.4 Baseline Blur Augmentations

Baseline comparisons use the out-of-the-box Albumentations [Buslaev et al., 2020] implementations without modification. Motion Blur is applied with a kernel size drawn uniformly from $\{3, 7\}$ at probability $p = 0.5$. Gaussian Blur uses a σ sampled uniformly from $[0.1, 2.0]$ at probability $p = 0.5$. Both are inserted at the same position in the pipeline as AuGhostmentation to ensure a fair comparison (Figure 8).

D.5 Modern Augmentations and Additional Architectures

Augmentation baselines use the ResNet-18 recipe above over five seeds, with AuGhostmentation prepended as a frontend transform in the combinations. ConvNeXt and ViT use AdamW with cosine decay and warmup, under a standard recipe and a DeiT-style recipe adding RandAugment, MixUp, CutMix, Random Erasing and label smoothing; on Tiny ImageNet they use the small-image stem of the ResNet-18 above, and ViT is trained from scratch or fine-tuned from ImageNet-1K weights. ConvNeXt, ViT and ResNeXt use five seeds, Swin-T and MaxViT-T three. A Tiny ImageNet run with ResNet-18 takes 540 s without augmentation, 1216 s with AuGhostmentation, 1777 s with AugMix,

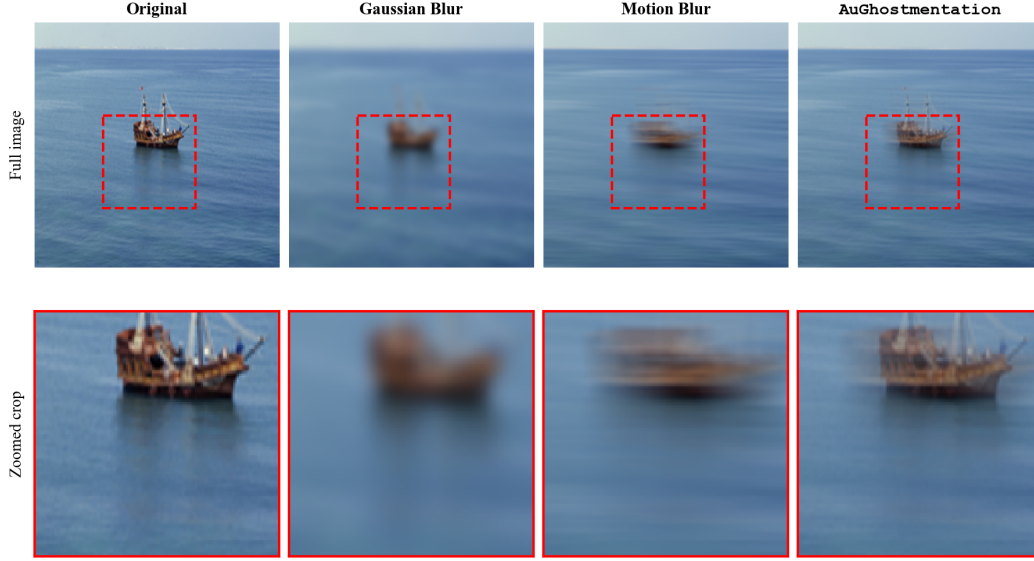


Figure 8: **Spatial artifacts introduced by different blur augmentations.** Each column shows the full image (top) and a zoomed crop of the highlighted region (bottom). Gaussian Blur uniformly smooths the scene, Motion Blur introduces directional streaking, while AuGhostmentation preserves local structure through a weighted ghost blend, producing a distinctly different artifact pattern.

Table 9: Tiny ImageNet accuracy (%) of ResNet-18 ($n = 5$) with each policy alone and with AuGhostmentation prepended.

Policy	Clean		mCA		
	Alone	+ AuGh.	Alone	+ AuGh.	Δ
Baseline	58.40	57.41	22.02	36.48	+13.56
AutoAugment	58.43	57.43	30.30	37.93	+7.63
AugMix	58.16	57.38	32.22	37.31	+5.09
RandAugment	59.69	59.08	28.06	39.02	+10.96
TrivialAugmentWide	58.59	58.85	33.83	39.13	+5.30
AugMix-JSD	59.51	56.99	36.54	40.02	+3.48
PixMix	59.07	57.14	24.44	35.15	+10.71

3260 s with AugMix-JSD and 4617 s with PixMix. Under the standard recipe, AuGhostmentation raises mCA by +3.55 for ViT trained from scratch and by +7.81 for ViT fine-tuned from ImageNet-1K weights. Clean accuracy moves from 27.61% to 28.64% for ViT from scratch and from 83.29% to 82.60% for fine-tuned ViT (DeiT-style), from 76.96% to 77.64% for Swin-T (standard), and from 75.66% to 76.87% for ConvNeXt-T on ImageNet-100. On Tiny ImageNet, MixUp [Zhang et al., 2018], CutMix [Yun et al., 2019] and Random Erasing [Zhong et al., 2020] alone reach 24.06%, 21.18% and 23.00% mCA (58.51%, 59.78% and 59.67% clean), close to the unaugmented 22.02%, and with ResNeXt [Xie et al., 2017] AuGhostmentation raises mCA from 23.88% to 36.42%. Under the standard recipe, shape bias rises from 80.6% to 83.4% for MaxViT-T and from 60.1% to 65.3% for Swin-T. Table 9 lists every policy with AuGhostmentation prepended; PixMix [Hendrycks et al., 2022] and AugMix-JSD modify the training objective.

E Stylized Robustness and Evaluation of Shape-bias

E.1 AdaIN Stylized Evaluation (Primary)

We evaluate all ResNet-50 checkpoints on the full ImageNet-100 validation set (5,000 images, 100 classes) as content, paired with a pool of 5,000 WikiArt paintings as styles (3 styles per content image, seed fixed to 0). Style transfer uses AdaIN [Huang and Belongie, 2017] with the pretrained VGG-19 encoder/decoder of Inoue [2019], swept over $\alpha \in \{0.25, 0.50, 0.75, 1.00\}$:

$$\mathbf{t} = \alpha \text{AdaIN}(\mathbf{f}_c, \mathbf{f}_s) + (1-\alpha) \mathbf{f}_c.$$

Content labels are preserved throughout (Figure 9). We report (i) **stylized top-1 accuracy**, (ii) **style-flip rate** $P(\hat{y}_{\text{styl}} \neq \hat{y}_{\text{clean}})$, and (iii) **matched stylized top-1**: accuracy restricted to trials where the clean prediction was already correct, isolating shape retention from baseline accuracy differences.

E.2 Cue-Conflict Shape-Bias Analysis (Confirmatory)

Inspired by Geirhos et al. [2019], we construct cue-conflict images by stylizing each content image (class A) with a donor drawn uniformly from a different class ($B \neq A$) at $\alpha = 1.0$. Each prediction is scored as a *shape win* ($\hat{y} = A$), *texture win* ($\hat{y} = B$), or *neither*; the shape-bias index follows Geirhos et al. [2019]:

$$\text{SB} = \frac{N_{\text{shape}}}{N_{\text{shape}} + N_{\text{texture}}}.$$

We repeat over 5 random seeds (5,000 pairs each) and report the mean SB with 95% bootstrap CIs. Forced-choice coverage (predictions in $\{A, B\}$) ranges from 21–28% across models and is reported alongside every estimate. This is an adaptation to the ImageNet-100 of the Geirhos et al. [2019] protocol. style donors are ImageNet-100 validation images rather than Painter-by-Numbers paintings, and predictions are scored at the 100-class level without a human legibility check; absolute SB values are therefore not directly comparable to those in Geirhos et al. [2019]. The Spearman rank correlation between WikiArt stylized top-1 (at $\alpha = 1.0$) and SB is $\rho = 1.0$ ($p < 0.001$) across all checkpoints, confirming that the WikiArt metric is a reliable proxy for shape-reliant behaviour in this model family. Under the same protocol, AugMix, AutoAugment, RandAugment and TrivialAugmentWide reach a shape bias of 77.0%, 76.2%, 76.0% and 74.2%.

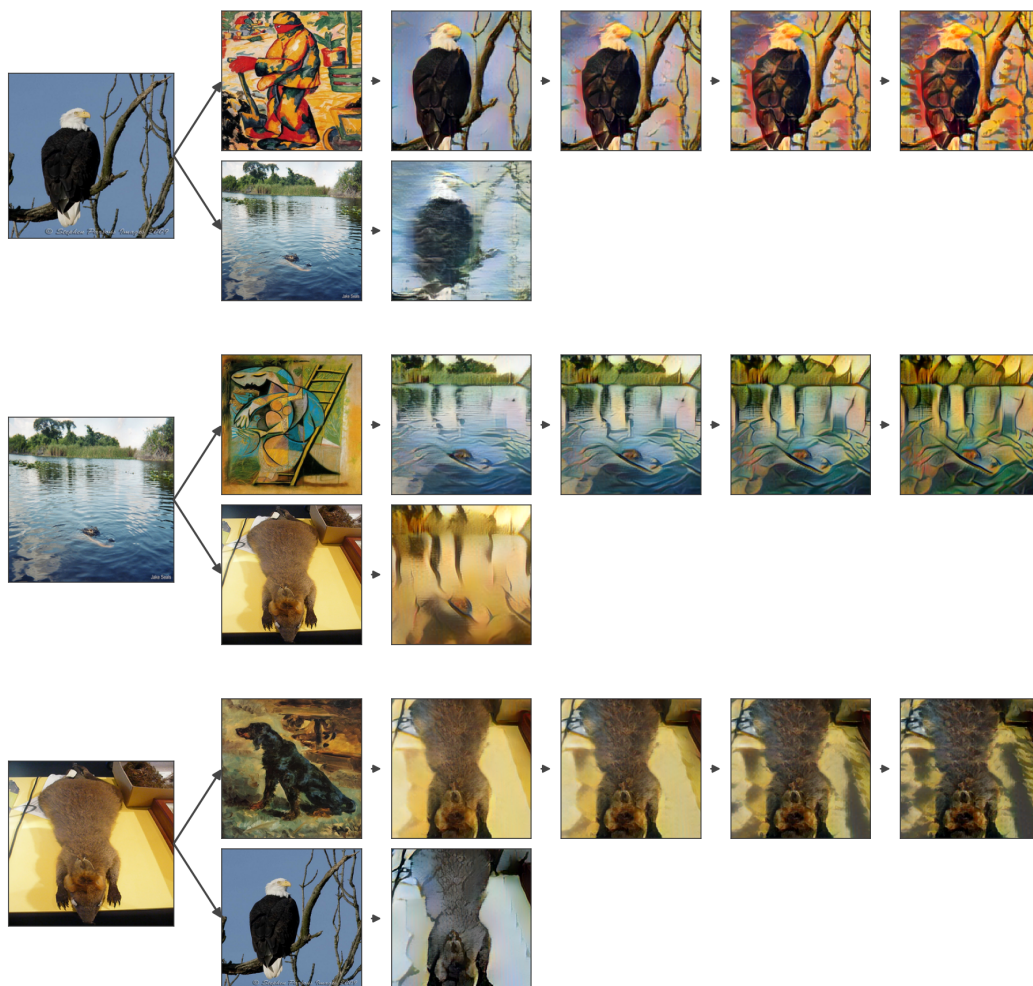


Figure 9: AuGhostmentation is tested both on painting-style transfer and cue-shape bias.

F Magnocellular Resemblance

The magnocellular (M) pathway—characterized by high temporal sensitivity, broad spatial tuning, and near-total exclusion of chromatic information—is disproportionately suppressed during saccades relative to the parvocellular pathway [Binda and Morrone, 2018, Ross et al., 2001]. This selective attenuation is consistent with the M-pathway’s primary role in motion and luminance-contrast processing: intra-saccadic streaks are luminance-defined directional signals, and their suppression is mediated by the same pathway that would otherwise relay them. AuGhostmentation’s directional blur kernels are, by construction, achromatic luminance-contrast signals—precisely the class of input to which the M-pathway is most sensitive.

This correspondence yields a testable prediction: removing chromatic content from training images should amplify the relative robustness gain, since the remaining signal is exclusively M-pathway in character. Reducing the three RGB channels to a single luminance channel (Figure 10) is broadly consistent with the prediction: AuGhostmentation achieves an 86.66% relative improvement in corruption robustness on grayscale Tiny ImageNet-C while preserving 98% of clean accuracy, compared to 66% and 98% respectively under standard color. However, Gaussian Blur—itsself an achromatic perturbation and therefore equally M-pathway-compatible—shows a comparable amplification under grayscale, indicating that the effect is not specific to AuGhostmentation’s directional structure. The more parsimonious account is that chromatic deprivation removes a shortcut the network would otherwise exploit, forcing reliance on luminance-defined geometric structure; all achromatic augmentations benefit from this shift, and the grayscale result therefore cannot be taken as evidence for the M-pathway hypothesis specifically [Nicolas et al., 2021].

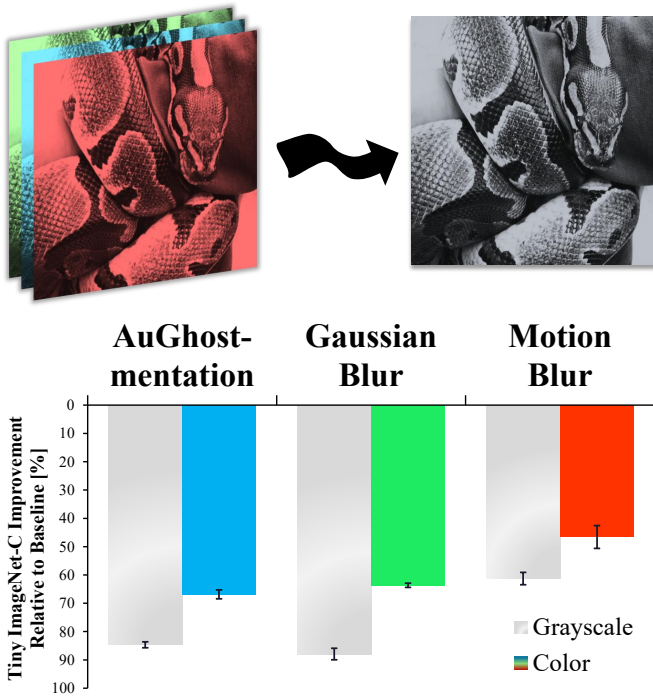


Figure 10: **ResNet-18 models trained and evaluated on both color and grayscale variants of Tiny ImageNet.**

Top: representative example of the color-to-grayscale conversion applied to test images. Bottom: Top-1 accuracy on clean Tiny ImageNet/Grayscale Tiny ImageNet and Tiny ImageNet-C/Grayscale Tiny ImageNet-C test sets across four training augmentation strategies. Bars show the mean over three random seeds; error bars denote standard deviation. Augmentation strategies improve robustness on corrupted inputs in both modalities, with larger relative gains under grayscale conditions.

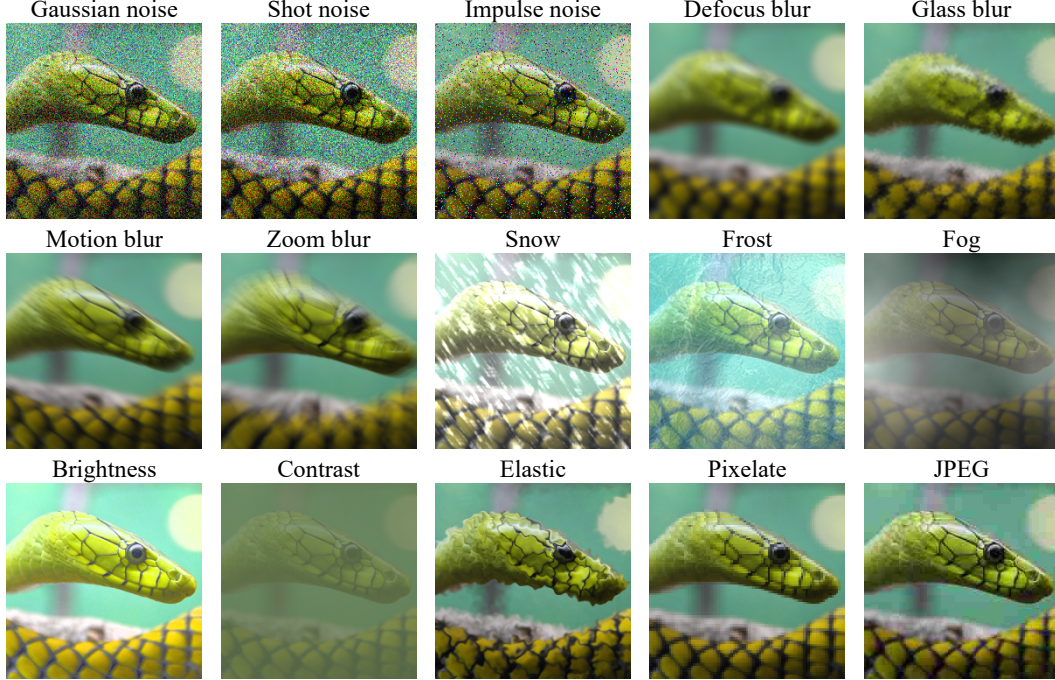


Figure 11: **The 15 ImageNet-C corruptions at severity 3**, applied to an ImageNet-100 validation image (green mamba) in the panel order of Figure 12.

Table 10: Resolution check at 160 px; the rule predicts $\lambda = 0.815$.

λ	Clean (%)	Test mCA (%)	Held-out mCA (%)
0.66	70.66	43.32	41.77
0.815	72.72	43.91	42.01
0.93	68.56	40.64	39.45
1.33	69.84	41.83	40.70
1.93	67.06	41.26	40.10

G Accuracy Across Severities

Figure 11 shows the 15 ImageNet-C corruptions on one validation image. Against Gaussian Blur, AuGhostmentation raises mCA by +0.66 on Tiny ImageNet and +1.7 on ImageNet-100, where the largest per-corruption gains are on fog (+14.5), motion blur (+10.6) and contrast (+7.1), and Gaussian Blur remains stronger on noise (28.5% vs. 25.1%). Thus, the gains of AuGhostmentation extend well beyond blur-like corruptions.

H Calibration of λ Across Resolutions

Formalizing the trend as $\lambda(r) = 0.33 + \log_2(224/r)$, where $\lambda = 0.33$ at 224px corresponds to $T \approx 330$ ms, predicts $\lambda = 0.815$ at 160 px, a resolution not used for selection. There it gave the best clean accuracy (72.72%), test mCA (43.91%) and held-out mCA (42.01%) among the evaluated values (Table 10). At 64 px it predicts $\lambda \approx 2.14$, against a best evaluated value of 1.93 on downsampled ImageNet; Tiny ImageNet, collected through a different pipeline, peaks at 1.33.

I Directional Prior: Statistics

Over the 21 values of p_h (five seeds each), one-way ANOVAs reject equal means for clean accuracy and mCA, upright and rotated, and the Spearman correlation between p_h and performance changes sign with rotation (Table 11); Kruskal–Wallis tests agree ($p < 0.015$ in all four conditions). With

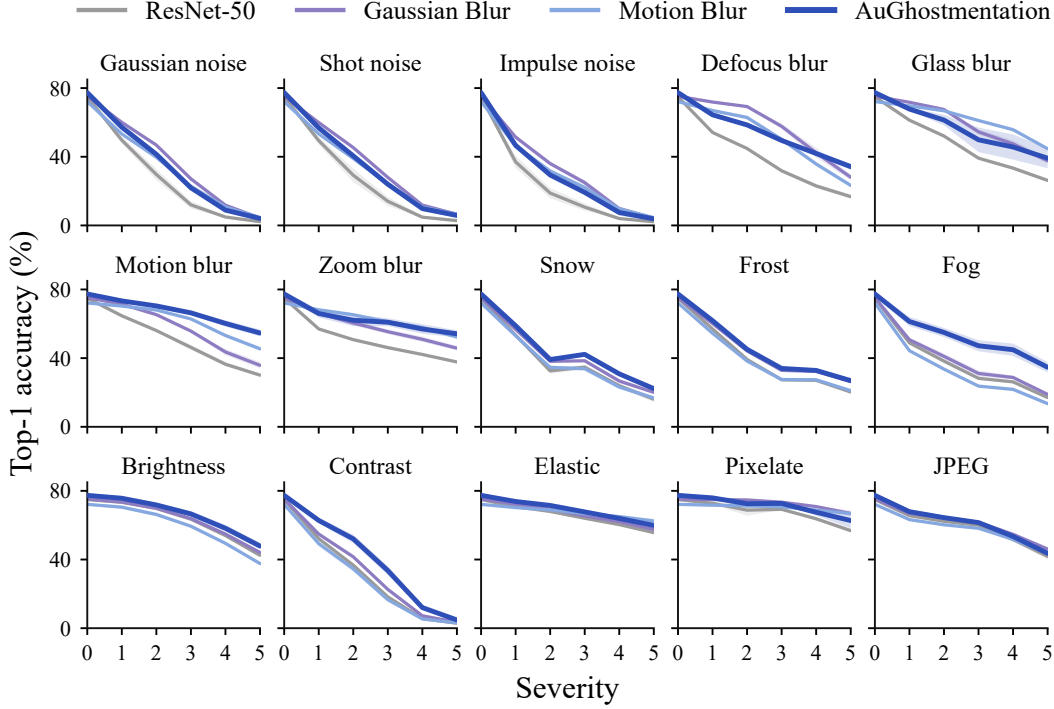


Figure 12: **ImageNet-100-C accuracy against corruption severity (ResNet-50)**. Panels follow the noise, blur, weather and digital families, and severity 0 is clean accuracy. Lines are means over three seeds.

Table 11: One-way ANOVA across p_h and Spearman correlation between p_h and performance.

Orientation	Metric	$F(20, 84), p$	Spearman ρ, p
Upright	Clean	1.88, 0.025	+0.53, 0.013
Upright	mCA	8.07, $< 10^{-6}$	+0.51, 0.018
Rotated 90°	Clean	1.94, 0.019	-0.61, 0.003
Rotated 90°	mCA	7.91, $< 10^{-6}$	-0.70, < 0.001

phase-scrambled images the preference is strongly attenuated (mean mCA 17.00% below and 17.21% above $p_h = 0.50$, peak 17.76% at 0.50), though scrambling also lowers overall accuracy. On ImageNet-100, $p_h = 0.99$ against 0.01 changes clean accuracy by +0.8 and mCA by +3.9 points. Phase scrambling lowers clean accuracy from 54% to 39% and mCA from 34% to 17%.

J Additional Ablations

The mean-matched control draws $M = 1$ with probability 0.67 and $M = 2$ with probability 0.33, so that $\mathbb{E}[M] = 1.33$ as under the Poisson model. It never leaves an image unaugmented, whereas the Poisson model leaves $e^{-1.33} = 26.4\%$ of images untouched. Applying the sampled kernels independently to I_0 and averaging, instead of compounding them, barely changes mCA (36.23% vs. 36.48%) but shifts it from blur (-1.77) to noise (+1.60), matching Gaussian Blur on the blur family (42.60% vs. 42.52%).

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope?

Answer: [Yes]

Justification: The abstract states three quantitative claims (60% relative robustness gain on Tiny ImageNet-C, 25% on VOneNet, +23% mCA and +4% clean accuracy on ImageNet-100); all are substantiated by Table 1 and Figure 4. Aspirational framing in the Introduction is clearly marked as motivation, not result.

Guidelines:

- The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The Conclusion names untested phenomena (luminance reduction during saccades, drift correction, predictive coding) and the Limitations section (Section ??) acknowledges the biological approximation may be imperfect. Results are scoped to ResNet, VGG, VOneNet, ResNeXt, ConvNeXt, ViT, Swin and MaxViT backbones on two subset benchmarks; generalization to larger architectures or full ImageNet is not claimed.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate “Limitations” section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren’t acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The paper presents no formal theorems or lemmas; all contributions are empirical. Mathematical derivations in the Appendix (kernel construction, MLE fitting) are engineering specifications, not proofs requiring theoretical assumptions.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The augmentation is fully specified in the Appendix (kernel construction, blending rule, and the fitted parameters p_h , κ , σ and median amplitude with their derivation from GazeBase v2.0), and the values of λ are given in Section 4.4. Training follows the cited recipes with the settings in Appendix D, and evaluation follows the standard ImageNet-C/Tiny ImageNet-C protocol (15 corruptions $\times$ 5 severities).

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The AuGhostmentation implementation is released under the MIT License at <https://github.com/emirhaninan/AuGhostmentation>, with usage instructions and the settings used for Tiny ImageNet and ImageNet-100. Training follows the publicly available recipes of Dapello et al. [2020] and Baidya et al. [2021] (Appendix D). All datasets are publicly available: GazeBase v2.0 (CC-BY-4.0), Tiny ImageNet, ImageNet-100 (a subset of ImageNet), ImageNet-C and Tiny ImageNet-C.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://neurips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://neurips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: Backbone architectures (ResNet-18 for Tiny ImageNet, ResNet-50 for ImageNet-100), evaluation benchmarks, number of independent seeds, and error-bar definition (standard deviation) are stated in the main text; full training hyperparameters appear in the supplemental material.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: All tables report mean $\pm$ standard deviation over 3–5 independent seeds; figure captions state this explicitly. Bootstrap confidence intervals ($B = 200$) are reported for fitted directional parameters in the Oculomotor Parameter Derivation appendix.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Experiments were run on Google Colab with opportunistic GPU allocation (NVIDIA RTX PRO 6000 Blackwell, A100 or L4, depending on availability) and locally on an NVIDIA RTX 4070 for select ablation runs; no single fixed hardware configuration was used throughout. Per-run wall-clock times for Tiny ImageNet/ResNet-18, measured on an RTX PRO 6000 Blackwell, are reported in Appendix D.5.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The work concerns image-classification augmentation using a publicly available eye-tracking dataset (GazeBase v2.0, CC-BY-4.0) collected under informed consent by its original authors; no new human data were collected and no sensitive or harmful content is involved.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: The paper presents a training augmentation that improves corruption robustness of image classifiers. There is no direct path to harmful applications; the method raises no fairness, privacy, or security concerns specific to this domain.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper releases no generative models, scraped datasets, or high-risk pre-trained artefacts; no safeguards are required.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [\[Yes\]](#)

Justification: GazeBase v2.0 is used under CC-BY-4.0 (stated in the Dataset appendix); VOneNet, AdaIN, ImageNet-100, and Tiny ImageNet are cited to their original papers. ImageNet is used under its standard non-commercial research licence.

Guidelines:

- The answer [\[N/A\]](#) means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [\[Yes\]](#)

Justification: The AuGhostmentation transform is the primary new asset; it is fully documented in the Formulation and Oculomotor Parameter Derivation appendices. The repository is anonymised at submission and will be released upon acceptance.

Guidelines:

- The answer [\[N/A\]](#) means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [\[N/A\]](#)

Justification: No human subjects research or crowdsourcing was conducted. GazeBase v2.0 is a pre-existing, publicly released dataset collected under informed consent by its original authors.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: No new human subjects research was conducted; IRB approval was not required.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs played no role in the core methodology. All biological parameters were derived computationally from GazeBase v2.0 using the Engbert–Kliegl algorithm and standard MLE fitting.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.