CORRUPTION ROBUSTNESS IN POST-TRAINING QUANTIZATION OF VISION TRANSFORMERS: A CALIBRATION PERSPECTIVE

Anonymous authors
Paper under double-blind review

ABSTRACT

Post-training quantization (PTQ) enables efficient deployment of pretrained models using only a small calibration set, but its robustness under distribution shifts remains underexplored. We observe that quantized models retain substantially less of their full-precision performance on corrupted inputs than on clean data, particularly under low-bit quantization. We study this problem from the perspective of calibration data and show that different transformations of the same fixed-budget calibration set lead to markedly different corruption robustness. Our analysis reveals that effective transformations tend to reduce activation energy, exhibiting a characteristic energy shift similar to that observed under common corruptions. We associate this reduction with two changes in self-attention: lower attention concentration and weaker directional alignment among value vectors associated with different tokens. Motivated by these observations, we introduce *spatial entropy of difference energy*, a model-independent score that measures how broadly a transformation distributes its changes across an image. This score strongly correlates with both activation-energy reduction and post-quantization corruption robustness, enabling transformation selection without model inference. Based on this proxy, we propose **Entropy-K**, which selects multiple high-scoring transformations while preserving the original calibration budget. Entropy-K requires neither corrupted target-domain samples nor modification of the underlying PTQ algorithm. Experiments across multiple Vision Transformer architectures and low-bit PTQ settings on ImageNet and ImageNet-C demonstrate consistent improvements in corruption robustness while maintaining competitive clean accuracy.

1 INTRODUCTION

Model quantization (Jacob et al., 2018) reduces memory and computational cost by representing model parameters and activations at lower numerical precision. Post-training quantization (PTQ) enables efficient deployment of pretrained models using only a small calibration set, without requiring model retraining. Accordingly, PTQ has been actively studied for the efficient deployment of modern Vision Transformers (ViTs) (Dosovitskiy et al., 2021; Touvron et al., 2021). Recent PTQ methods for ViTs have substantially improved low-bit quantization performance through specialized quantizers (Yuan et al., 2022; Li et al., 2023; Liu et al., 2023), rounding optimization (Nagel et al., 2020), block reconstruction (Li et al., 2021; Wei et al., 2022), and sensitivity-aware reconstruction (Wu et al., 2025b;a; Hwang et al., 2026). However, most existing methods primarily focus on preserving in-distribution (ID) performance, while robustness to distribution shifts caused by common corruptions, such as noise and blur, has received limited attention despite its importance for reliable deployment. Motivated by this limitation, we study corruption robustness in low-bit PTQ for ViTs.

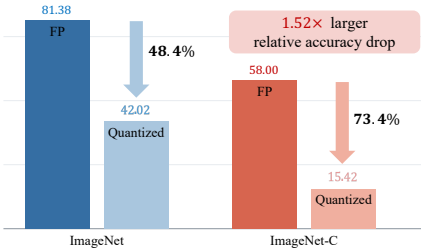


Figure 1: **Relative accuracy preservation after quantization.** The relative accuracy drop is substantially larger on ImageNet-C than on ImageNet.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We observe that quantization preserves substantially less of the full-precision (FP) model's performance on corrupted ImageNet-C (Hendrycks & Dietterich, 2019) inputs than on clean ImageNet (Deng et al., 2009) inputs, as illustrated in Figure 1. We study this problem from the perspective of calibration data, since the calibration samples determine the FP activations used to optimize the quantization parameters. A straightforward approach is to apply existing augmentation methods (Yun et al., 2019; Hendrycks et al., 2020; Modas et al., 2022) to diversify the calibration samples; however, adding augmented samples increases the calibration-set size and reconstruction cost. To preserve the efficiency of PTQ, we instead transform the given calibration samples without increasing the calibration-set size. Interestingly, different transformations lead to markedly different corruption robustness, with some improving robustness and others providing little benefit or even degrading it. This observation suggests that post-quantization robustness depends strongly on how the calibration samples are transformed.

To understand what makes a transformation effective, we analyze the FP activations induced by different calibration transformations. We find that transformations yielding larger corruption-robustness gains tend to reduce activation energy, exhibiting a similar energy shift to that observed under common corruptions. This activation behavior provides a useful indicator of transformation effectiveness, but directly measuring it requires repeated FP inference and activation extraction for every candidate transformation. Such model-dependent evaluation becomes costly when transformation selection must be repeated across many candidate transformations, model architectures, or quantization settings. To avoid this overhead, we seek a forward-free indicator that can be computed directly from the calibration images. Accordingly, we introduce the *spatial entropy of difference energy*, $S_{\mathrm{ent}}(T)$, which measures how broadly a transformation distributes its differences across an image. We find that higher S_{ent} is strongly associated with both activation-energy reduction and improved post-quantization corruption robustness. Based on this score, we propose **Entropy-K**, which selects multiple high-scoring transformations while preserving the original calibration budget. Experiments across seven Vision Transformer architectures, multiple low-bit settings, and different PTQ methods demonstrate consistent improvements in corruption robustness while maintaining competitive ID accuracy.

Our main contributions are summarized as follows:

- We reveal that low-bit PTQ for Vision Transformers preserves substantially less FP performance under common corruptions than on clean inputs, particularly at lower bit-widths.
- We identify activation-energy reduction as a common characteristic of effective calibration transformations and corrupted inputs, and explain it through changes in self-attention.
- We introduce S_{ent}, a forward-free and model-independent indicator that captures the spatial distribution of transformation-induced differences and is strongly associated with both activation-energy reduction and corruption robustness.
- Building on this indicator, we propose Entropy-K, a fixed-budget calibration strategy that improves corruption robustness without additional training data, target-domain samples, model retraining, or modification of the underlying PTQ algorithm.

2 Related Work

Post-training quantization. Existing PTQ methods have improved low-bit quantization through rounding and reconstruction optimization (Nagel et al., 2020; Li et al., 2021; Wei et al., 2022). For ViTs, PTQ4ViT (Yuan et al., 2022), RepQ-ViT (Li et al., 2023), NoisyQuant (Liu et al., 2023), and I&S-ViT (Zhong et al., 2026) address challenging activation distributions through specialized quantization, reparameterization, and optimization strategies. More recent methods, including APHQ-ViT (Wu et al., 2025b), FIMA-Q (Wu et al., 2025a), and LS-ViT (Hwang et al., 2026), further improve reconstruction by incorporating sensitivity information. Despite these advances, existing methods primarily focus on the quantization or reconstruction procedure while relying on a small set of clean calibration samples. Unlike these approaches, our work focuses on the role of calibration data rather than modifying the quantization or reconstruction procedure.

Calibration data and quantization robustness. Recent studies have examined the role of calibration data in PTQ. SelectQ (Zhang et al., 2025) selects informative calibration samples, but selection-based approaches require flexibility in constructing the calibration set and are less applicable when

the calibration samples are fixed or predefined. MetaAug (Pham et al., 2024) takes a different direction by learning calibration transformations through bi-level optimization, requiring an additional optimization stage. These studies demonstrate the importance of calibration data, but require either sample selection or an additional learning procedure. Separately, RobustMQ (Xiao et al., 2023) reports the increased vulnerability of quantized models to natural corruptions, while Recti-Q (Araghi et al., 2026) improves post-quantization robustness by introducing a lightweight feature adapter. Our work instead studies corruption robustness from the calibration-data perspective by analyzing how predefined transformations of a fixed calibration set affect FP activation behavior during reconstruction. Based on this analysis, we select effective transformations without additional learning or trainable parameters, allowing the proposed strategy to be directly applied to existing PTQ methods without modifying either the quantization procedure or the deployed model.

Data transformations and activation behavior. Data augmentation has been widely studied to improve model generalization through diverse modifications of training samples. General-purpose methods include RandomErasing (Zhong et al., 2020), MixUp (Zhang et al., 2018), CutMix (Yun et al., 2019), RandAugment (Cubuk et al., 2020), and TrivialAugment (Müller & Hutter, 2021), while AugMix (Hendrycks et al., 2020), PRIME (Modas et al., 2022), HybridAugment++ (Yucel et al., 2023), and RVT-PWA (Mao et al., 2022) place greater emphasis on robustness to common corruptions. These methods are primarily designed for model training, where transformed samples influence the learned model parameters through optimization. In contrast, under our PTQ setting, the FP model remains fixed, and such transformations alter the FP activations presented during reconstruction, thereby influencing the optimization of quantization parameters. Prior studies have shown that activation magnitude reflects representation changes under distribution shifts (Yu et al., 2023; Park et al., 2023), motivating our analysis of how calibration transformations alter activation energy and post-quantization robustness.

3 Preliminaries and Problem Formulation

Given a pretrained full-precision (FP) model Θ and a small calibration set $\mathcal{D}_{\text{cal}}$, post-training quantization (PTQ) determines quantization parameters without retraining the pretrained model. In conventional PTQ, $\mathcal{D}_{\text{cal}}$ is typically given or randomly sampled from the source distribution. Since calibration samples determine the FP activations used during quantization, we investigate whether transforming these samples can improve corruption robustness.

Rather than enlarging the calibration set, which increases the cost of reconstruction-based PTQ (Li et al., 2021; Hwang et al., 2026), we keep the calibration budget fixed. Given a calibration transformation T, we construct the transformed calibration set as

$$\mathcal{D}_{\text{cal}}^{T} = \{T(\boldsymbol{x}) \,|\, \boldsymbol{x} \in \mathcal{D}_{\text{cal}}\}, \qquad |\mathcal{D}_{\text{cal}}^{T}| = |\mathcal{D}_{\text{cal}}|. \tag{1}$$

Let $a_l^{\text{FP}}(\boldsymbol{x})$ and $a_l^{\text{Q}}(\boldsymbol{x};\theta_l)$ denote the FP and quantized activations at the l-th block, respectively. Using $\mathcal{D}_{\text{cal}}^{T}$, reconstruction-based PTQ determines the quantization parameters as

$$\theta_l^{T} = \arg\min_{\theta_l} \frac{1}{|\mathcal{D}_{\text{cal}}^{T}|} \sum_{\boldsymbol{x}\in\mathcal{D}_{\text{cal}}^{T}} d\Big(a_l^{\text{FP}}(\boldsymbol{x}), a_l^{\text{Q}}(\boldsymbol{x};\theta_l)\Big), \tag{2}$$

where $d(\cdot,\cdot)$ denotes the reconstruction discrepancy. Thus, different transformations induce different FP activations and can lead to different quantization parameters even under the same calibration budget and PTQ algorithm.

Let Θ_{Q}^{T} denote the quantized model obtained using $\mathcal{D}_{\text{cal}}^{T}$, and $\Delta_{\mathcal{D}}^{T}$ denote its performance degradation relative to the FP model Θ on distribution $\mathcal{D}$. Let $\mathcal{T}$ denote the set of candidate calibration transformations. To characterize a desirable calibration transformation, we seek $T \in \mathcal{T}$ that minimizes degradation on corrupted inputs while limiting additional degradation on ID data:

$$T^{*} = \arg\min_{T\in\mathcal{T}} \Delta_{\text{C}}^{T} \quad \text{s.t.} \quad \Delta_{\text{ID}}^{T} - \Delta_{\text{ID}}^{I} \le \epsilon, \tag{3}$$

where I denotes the identity transformation corresponding to conventional clean calibration, and ϵ controls the allowable additional ID degradation. Since corrupted target data are unavailable during calibration, Eq. 3 only characterizes the desired transformation, motivating a target-free criterion for identifying it using the given calibration set.

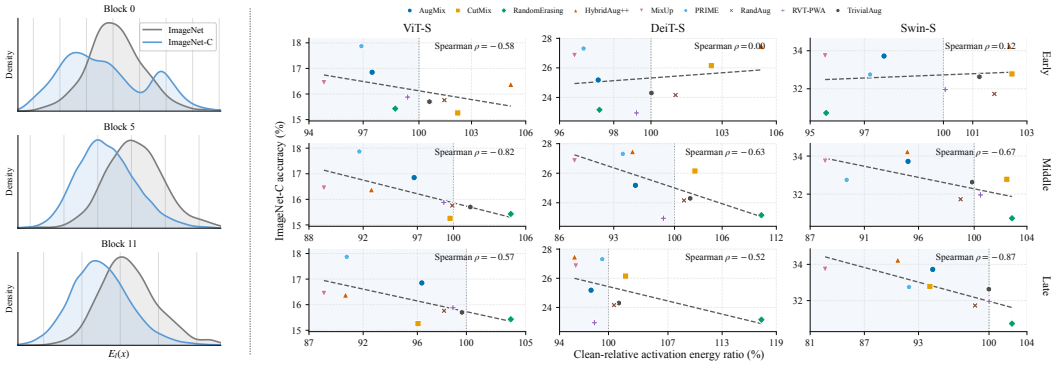


Figure 3: **Activation-energy analysis under corruption and calibration transformations. (Left)** FP activation-energy distributions of ImageNet and ImageNet-C inputs at Blocks 0, 5, and 11 of ViT-S. **(Right)** Relationship between the clean-relative activation-energy ratio induced by calibration transformations and ImageNet-C accuracy across different ViT architectures. The rows in the right panel correspond to representative early, middle, and late blocks, respectively.

4 Corruption Robustness Analysis

Which calibration transformations improve corruption robustness? We examine whether calibration transformations can improve corruption robustness after quantization. Following the fixed-budget setting defined in Section 3, we apply several representative transformations adopted from existing augmentation methods to the same calibration set. Figure 2 reports the change in ImageNet-C accuracy relative to conventional clean calibration under W2A3 quantization, where W2A3 denotes 2-bit weights and 3-bit activations. The results show that calibration transformations can have substantially different effects on corruption robustness. While some transformations improve ImageNet-C accuracy over clean calibration, others provide only limited gains or even degrade performance. This variation motivates us to investigate what properties distinguish effective transformations from less effective ones.

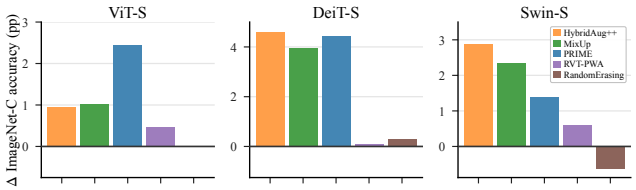


Figure 2: **Effect of calibration transformations on ImageNet-C accuracy.** The y-axis denotes the accuracy change relative to conventional clean calibration.

Activation analysis under corruption. We examine how common corruptions alter the FP activations compared with clean inputs. To characterize these changes, we analyze activation energy across transformer blocks, following prior studies that use activation magnitude as a signal for out-of-distribution detection (Yu et al., 2023; Park et al., 2023). For an input $\boldsymbol{x}$, we define the activation energy at the l-th block as

$$E_l(\boldsymbol{x}) = \left\| a_l^{\mathrm{FP}}(\boldsymbol{x}) \right\|_F^2, \tag{4}$$

where $a_l^{\mathrm{FP}}(\boldsymbol{x})$ denotes the FP activation at the l-th block. The left panel of Figure 3 compares the activation-energy distributions of clean ImageNet and corrupted ImageNet-C inputs in ViT-S. We visualize representative early, middle, and late blocks, corresponding to Blocks 0, 5, and 11, respectively. Compared with clean inputs, corrupted inputs generally produce lower activation energy. While the difference is less consistent in the early block, the corrupted distribution exhibits a clearer shift toward lower energy in the middle and late blocks.

Does transformation-induced energy reduction correlate with corruption robustness? We investigate whether activation-energy changes induced by calibration transformations are associated with ImageNet-C accuracy after quantization. We consider nine representative transformations: RandomErasing (Zhong et al., 2020), CutMix (Yun et al., 2019), MixUp (Zhang et al., 2018), RandAugment (Cubuk et al., 2020), TrivialAugment (Müller & Hutter, 2021), AugMix (Hendrycks et al., 2020), RVT-PWA (Mao et al., 2022), HybridAugment++ (Yucel et al., 2023),

and PRIME (Modas et al., 2022). Motivated by the block-dependent energy shifts observed in the left panel of Figure 3, we separately analyze representative early, middle, and late blocks. The right panel reports, for each transformation, the clean-relative activation-energy ratio against the resulting ImageNet-C accuracy, with columns corresponding to different architectures and rows to different network stages. An energy ratio below 100% indicates that the transformation reduces FP activation energy relative to clean calibration. In the early blocks, the relationship is weak and varies substantially across architectures. In contrast, the middle blocks exhibit a consistently strong negative association, indicating that transformations inducing lower activation energy tend to yield higher corruption accuracy. The late blocks show a similar negative trend, but with greater variation across architectures. These results identify middle-block activation-energy reduction as a consistent characteristic of effective calibration transformations, motivating the following analysis. We further investigate the role of activation magnitude through direct activation-scaling experiments in Appendix E.

Why do corrupted inputs yield lower activation energy? To understand the observed energy reduction, we analyze self-attention, a core component of modern ViTs. We find that corruption reduces attention-head output energy through two complementary effects: less concentrated attention weights and weaker directional alignment among value vectors. For a single attention head, let $A \in \mathbb{R}^{Q\times K}$ denote the attention matrix and $V \in \mathbb{R}^{K\times d}$ denote the value matrix, where Q and K are the numbers of query and key tokens, respectively. The analysis is applied independently to each head, and the resulting statistics are aggregated across heads and middle blocks. The attention-head output energy can be decomposed as

$$E_{\text{attn}} = \|AV\|_F^2 = \underbrace{\sum_{i=1}^{Q}\sum_{j=1}^{K} a_{ij}^2 \,\|\boldsymbol{v}_j\|_2^2}_{D\text{: diagonal energy}} + \underbrace{\sum_{i=1}^{Q}\sum_{\substack{j,k=1\\ j\neq k}}^{K} a_{ij}a_{ik}\,\langle \boldsymbol{v}_j, \boldsymbol{v}_k\rangle}_{X\text{: cross energy}}. \tag{5}$$

The diagonal term D collects the $j = k$ contributions and reflects individual value-vector energies weighted by squared attention weights. The cross term X collects the $j \neq k$ contributions and reflects interactions between different value vectors, increasing when their directions are more strongly aligned. The full derivation of Eq. 5 is provided in Appendix D.

We first examine the diagonal term by measuring how concentrated the attention weights are. We define the attention concentration as

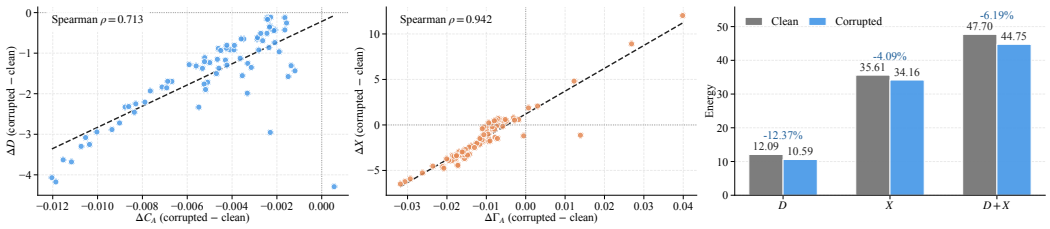
$$C_A = \frac{1}{Q}\sum_{i=1}^{Q}\sum_{j=1}^{K} a_{ij}^2. \tag{6}$$

For each query, the sum of squared attention weights becomes larger when attention is concentrated on a small number of key tokens and smaller when it is distributed more uniformly across tokens. This dispersion is also closely related to higher attention entropy. Accordingly, a lower C_A indicates more dispersed attention. The diagonal energy D depends on the same squared attention weights, with each term additionally scaled by the corresponding value-vector energy $\|\boldsymbol{v}_j\|_2^2$. When value-vector energies are comparable across tokens, D can be approximated as $D \approx Q\,\overline{\|\boldsymbol{v}\|_2^2}\,C_A$, where $\overline{\|\boldsymbol{v}\|_2^2}$ denotes the mean value-vector energy. Thus, more dispersed attention can reduce the diagonal energy even without a reduction in value-vector magnitude. As shown in Figure 4a, corrupted inputs consistently exhibit lower C_A than clean inputs. Across corruption–severity conditions, the reduction in C_A is strongly correlated with the reduction in D, with a Spearman correlation of $\rho = 0.713$. Consistently, Figure 4c shows that D decreases from 12.09 to 10.59, corresponding to a 12.37% reduction. These results support reduced attention concentration as a major factor associated with the decrease in diagonal energy under corruption.

We also examine the cross term X, focusing on the directional alignment among value vectors. To quantify this alignment, we define

$$\Gamma_A = \frac{X}{\sum_{i=1}^{Q}\sum_{\substack{j,k=1\\ j\neq k}}^{K} a_{ij}a_{ik}\|\boldsymbol{v}_j\|_2\|\boldsymbol{v}_k\|_2}. \tag{7}$$

By normalizing X with the corresponding attention-weight and magnitude terms, Γ_A is designed to capture the directional alignment among different value vectors while reducing the influence of



(a) Attention concentration (C_A). (b) Value alignment (Γ_A). (c) Attention energy ($D + X$).

Figure 4: **Corruption reduces attention-head output energy.** Each point in (a) and (b) represents one corruption–severity condition. Corrupted inputs exhibit lower attention concentration and weaker directional alignment among value vectors, which are associated with reductions in the diagonal energy D and cross energy X, respectively. Together, these effects reduce the total attention-head output energy.

their magnitudes and attention scales. A larger Γ_A indicates stronger directional alignment among value vectors, increasing their contribution to X, whereas a smaller Γ_A indicates weaker directional reinforcement. As shown in Figure 4b, corrupted inputs generally exhibit lower Γ_A than clean inputs, and the reduction in Γ_A is strongly correlated with the reduction in cross energy ($\rho = 0.942$). Accordingly, Figure 4c shows that the cross energy X decreases from 35.61 to 34.16, corresponding to a 4.09% reduction. These results indicate that corruption weakens the directional alignment among value vectors, contributing to the reduction in cross energy.

Taken together, corruption reduces attention-head output energy through two complementary effects: it disperses attention across more tokens, lowering the diagonal energy D, and weakens the directional alignment among value vectors, lowering the cross energy X. Combined, these effects are consistent with the observed 6.19% reduction in total attention-head output energy and help explain the lower activation energy under corrupted inputs.

5 Forward-Free Proxy for Transformation Selection

The middle-block activation energy provides a useful indicator of transformation effectiveness, but measuring it requires forwarding every transformed calibration sample through the FP model and extracting its activations. This cost scales with the number of candidate transformations, calibration samples, and model architectures. For example, evaluating M transformations on N calibration samples across A architectures requires $A \times M \times N$ FP forward passes. Such model-dependent evaluation becomes increasingly burdensome when transformation selection must be repeated across different architectures or quantization settings. We therefore seek an input-space proxy that can estimate transformation-induced energy reduction directly from clean and transformed images, without requiring either FP or quantized-model inference. Our analysis in Section 4 provides a useful clue for constructing such a proxy. Corrupted inputs exhibit lower attention concentration and weaker directional alignment among value vectors, both of which are associated with reduced attention-head output energy. These observations suggest that transformation effectiveness may depend not only on the magnitude of input changes, but also on how broadly those changes are distributed across spatial tokens. Motivated by this, we examine whether the spatial distribution of transformation-induced differences can serve as a forward-free proxy for the resulting activation-energy shift.

Spatial entropy of difference energy. We partition an input image into P fixed non-overlapping regions and measure the transformation-induced difference within each region. For the j-th region of an input $\boldsymbol{x}$, denoted by $\boldsymbol{x}_j$, we define the difference energy as

$$e_j(\boldsymbol{x}, T) = \|T(\boldsymbol{x})_j - \boldsymbol{x}_j\|_2^2 . \tag{8}$$

We then quantify how broadly these differences are distributed across spatial regions using the normalized entropy

$$H(\boldsymbol{x}, T) = -\frac{1}{\log P} \sum_{j=1}^{P} p_j(\boldsymbol{x}, T) \log p_j(\boldsymbol{x}, T), \qquad p_j(\boldsymbol{x}, T) = \frac{e_j(\boldsymbol{x}, T)}{\sum_{k=1}^{P} e_k(\boldsymbol{x}, T)}. \tag{9}$$

The normalization by $\log P$ bounds the entropy within $[0, 1]$. We assign a score to each transformation T by averaging the entropy over the calibration set:

$$S_{\text{ent}}(T) = \frac{1}{|\mathcal{D}_{\text{cal}}|} \sum_{\boldsymbol{x} \in \mathcal{D}_{\text{cal}}} H(\boldsymbol{x}, T). \tag{10}$$

When $S_{\text{ent}}(T)$ approaches zero, the difference energy is concentrated in a small number of spatial regions. In contrast, when $S_{\text{ent}}(T)$ approaches one, the difference energy is distributed more uniformly across the image. Thus, a larger $S_{\text{ent}}(T)$ indicates that the transformation induces more spatially distributed changes across the input.

We examine whether S_{ent} captures the activation-energy behavior associated with effective calibration transformations and serves as an indicator of corruption robustness. Figure 5 compares S_{ent} with the clean-relative middle-block activation energy and ImageNet-C accuracy gain across candidate transformations. Each marker represents a transformation and reports the mean over ViT-S, DeiT-S, and Swin-S, while the vertical bar indicates the range across the three architectures. As shown in the left panel, higher S_{ent} is strongly associated with lower middle-block activation energy relative to clean calibration, with a Spearman correlation of $\rho = -0.78$. This trend suggests that transformations inducing more spatially distributed differences tend to produce larger reductions in middle-block activation energy. More importantly, the right panel shows that higher S_{ent} is also associated with larger ImageNet-C accuracy gains, with $\rho = 0.83$. Although the magnitude of the gain varies across architectures, the overall trend remains consistent across transformations. Together, these results support S_{ent} as a useful forward-free indicator for transformation selection, as it is strongly associated with both middle-block activation-energy reduction and post-quantization corruption robustness. Further validation of S_{ent} is provided in Appendix B.

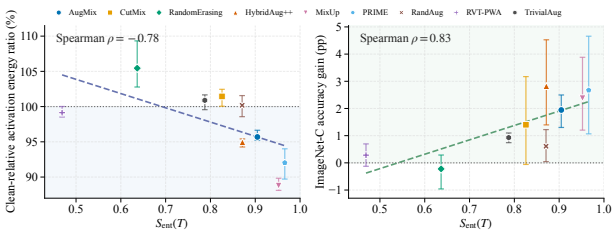


Figure 5: $S_{\text{ent}}(T)$ **correlates with (left) activation-energy reduction and (right) corruption-robustness gains.**

Entropy-guided transformation selection. Using S_{ent} as a forward-free proxy, we rank candidate transformations without model inference. Specifically, we rank the candidate set $\mathcal{T}$ according to $S_{\text{ent}}(T)$ and select the top-K transformations with the highest scores. As indicated by the architecture-wise variation in Figure 5, the effectiveness of a single transformation can still vary across models. We therefore use K high-scoring transformations rather than relying on a single candidate, reducing sensitivity to architecture-specific transformation effects. We refer to this strategy as **Entropy-K**. In all classification experiments, we set $K = 3$, selecting PRIME, AugMix, and MixUp as the top-ranked transformations. To preserve the fixed calibration budget, we randomly partition the calibration set into K disjoint subsets and apply one selected transformation to each subset, keeping the total number of calibration samples unchanged.

6 Experiments

Experimental setup. We evaluate our method on seven ImageNet-pretrained Vision Transformer architectures: ViT-S/B (Dosovitskiy et al., 2021), DeiT-T/S/B (Touvron et al., 2021), and Swin-S/B (Liu et al., 2021). For PTQ, we sample 1,024 ImageNet training images, shared across all experiments; additional calibration-budget results are provided in Appendix J. We evaluate in-distribution (ID) performance on ImageNet and corruption robustness (C) on ImageNet-C. Experiments are conducted under two low-bit quantization settings, W3A3 and W2A3, where W and A denote the bit-widths of weights and activations, respectively. To compute S_{ent}, each image is partitioned into fixed 16×16-pixel regions. Additional experimental details are provided in Appendix C.

Quantitative comparison. We compare Entropy-K with representative PTQ methods, including APHQ-ViT (Wu et al., 2025b), FIMA-Q (Wu et al., 2025a), and LS-ViT (Hwang et al., 2026). We use LS-ViT as the primary PTQ baseline and further evaluate nine representative calibration transformations applied on top of it: RandomErasing, CutMix, MixUp, RandAugment, TrivialAugment, AugMix, RVT-PWA, HybridAugment++, and PRIME. As shown in Table 1, the effectiveness of individual transformations varies substantially across model architectures and quantization settings. A

Method	W/A	ViT-S		ViT-B		DeiT-T		DeiT-S		DeiT-B		Swin-S		Swin-B		Avg.
		ID	C	ID	C	ID	C	ID	C	ID	C	ID	C	ID	C	
Full-Prec	32/32	81.38	58.00	85.10	69.22	72.17	43.76	79.86	56.30	81.98	61.15	83.32	63.34	85.27	67.02	70.56
APHQ-ViT	3/3	64.37	31.26	73.67	41.81	56.08	25.78	68.38	37.28	76.15	49.01	77.35	44.29	77.95	50.02	55.24
FIMA-Q	3/3	63.50	29.99	74.54	43.69	55.81	25.72	68.68	37.56	76.27	48.83	77.27	43.27	78.67	48.72	55.18
LS-ViT	3/3	64.50	31.17	74.92	45.10	55.86	25.41	69.18	38.11	76.48	49.20	77.91	44.76	79.50	49.85	55.85
+RandomErasing	3/3	64.07	30.91	74.82	44.76	55.33	25.34	69.07	38.25	76.08	48.02	77.55	43.66	79.10	49.37	55.45
+CutMix	3/3	63.88	30.98	75.30	44.61	57.09	26.56	70.73	39.73	77.00	48.87	78.80	45.24	80.04	51.03	56.42
+MixUp	3/3	63.55	32.46	74.39	45.99	55.64	27.24	69.75	40.17	76.25	50.59	78.40	46.78	78.97	50.76	56.50
+RandAugment	3/3	64.54	31.94	74.91	45.43	55.70	25.59	69.24	38.91	76.33	49.60	77.64	44.38	79.14	50.48	55.99
+TrivialAugment	3/3	64.58	32.04	74.94	45.81	55.81	26.18	69.32	38.84	76.43	49.42	78.01	45.32	79.46	51.03	56.23
+AugMix	3/3	64.41	32.49	74.85	46.48	56.15	26.66	69.76	39.74	76.42	50.14	78.29	45.94	79.31	50.73	56.53
+RVT-PWA	3/3	64.55	31.28	74.96	44.91	56.00	25.75	68.89	38.02	76.47	49.08	77.89	44.66	79.38	49.83	55.83
+HybridAugment++	3/3	63.47	32.27	74.53	46.72	56.21	26.91	70.13	41.08	76.64	50.59	78.13	46.57	79.37	51.33	56.71
+PRIME	3/3	63.34	33.41	74.67	48.25	55.71	27.74	69.54	41.41	76.30	51.11	77.72	46.28	79.15	50.80	<u>56.82</u>
+**Entropy-K**	3/3	64.65	33.82	74.92	48.49	56.32	27.71	69.93	40.94	76.75	50.84	78.24	46.85	79.30	51.15	**57.14**
APHQ-ViT	2/3	43.07	15.58	53.12	20.22	37.83	13.40	50.33	20.68	66.19	34.31	65.95	29.89	64.44	31.72	39.05
FIMA-Q	2/3	40.72	14.82	55.61	22.66	37.90	13.97	51.21	22.37	66.46	34.96	64.88	28.33	63.70	28.99	39.04
LS-ViT	2/3	42.02	15.42	56.76	23.43	38.63	14.26	52.45	22.87	67.00	35.73	67.73	31.37	67.76	34.16	40.69
+RandomErasing	2/3	42.54	15.43	56.47	23.61	36.33	12.80	53.37	23.16	66.44	33.63	67.38	30.73	66.49	32.05	40.03
+CutMix	2/3	43.04	15.26	57.78	23.82	39.73	14.34	56.84	26.15	68.16	35.05	69.72	32.78	69.69	34.05	41.89
+MixUp	2/3	41.95	16.44	56.73	26.26	37.91	15.85	55.19	26.84	66.66	38.10	68.77	33.73	65.89	32.32	41.62
+RandAugment	2/3	41.18	15.76	57.37	24.65	37.98	14.43	53.18	24.16	67.01	36.37	67.34	31.73	68.20	34.45	40.99
+TrivialAugment	2/3	42.06	15.70	55.94	24.10	37.64	14.11	53.53	24.30	67.14	36.43	68.52	32.63	67.68	34.52	41.02
+AugMix	2/3	43.73	16.85	57.17	25.61	38.77	14.91	54.35	25.18	67.30	37.49	68.90	33.72	68.67	34.75	<u>41.96</u>
+RVT-PWA	2/3	43.26	15.88	57.28	24.17	38.20	14.07	52.99	22.95	66.94	35.98	67.39	31.96	68.24	33.83	40.94
+HybridAugment++	2/3	40.40	16.38	55.87	25.18	38.23	14.95	55.61	27.47	67.11	38.58	68.99	34.24	67.81	35.76	41.90
+PRIME	2/3	40.92	17.87	56.65	27.68	37.38	16.07	53.89	27.31	66.83	38.65	67.26	32.75	65.56	33.47	41.59
+**Entropy-K**	2/3	42.47	17.78	57.32	27.83	38.40	16.07	54.76	26.98	67.25	39.24	68.36	34.15	67.72	34.52	**42.35**

Table 1: **Quantization performance on ImageNet and ImageNet-C.** ID and C denote ImageNet and ImageNet-C accuracy, respectively, and Avg. is the mean over all 14 ID and C results across seven architectures. For transformation-based methods, red and blue indicate changes of at least ± 1.0 percentage point relative to LS-ViT under the same quantization setting. The best and second-best Avg. results are shown in **bold** and <u>underline</u>, respectively.

transformation that improves corruption robustness for one configuration can provide only marginal gains or even degrade performance for another. Moreover, several transformations exhibit a trade-off between ID accuracy and corruption robustness, indicating that improved corruption robustness does not necessarily preserve ID performance. In contrast, Entropy-K maintains a more favorable balance between ID and corruption performance by combining multiple high-scoring transformations rather than relying on a single predefined transformation. Entropy-K achieves consistent gains under both W3A3 and the more aggressive W2A3 setting, indicating that the proposed transformation selection remains effective across different quantization severities. Overall, these results show that Entropy-K provides more consistent performance across architectures, bit-widths, and evaluation distributions. Additional PTQ baselines and W2A4 results are provided in Appendix K.

Effect of transformation selection. We analyze the effect of the number of selected transformations under W3A3 quantization on ViT-S. Table 2(a) compares Entropy-K with different values of K, where Entropy-9 uses all candidate transformations. Using a single transformation improves corruption accuracy over the clean-calibration baseline, but incurs ID accuracy degradation. Increasing K alleviates this trade-off by combining multiple transformations, leading to better preservation of ID performance while retaining the robustness gain. In particular, $K = 3$ achieves the highest corruption accuracy and the best average performance. Increasing K to 9 provides no additional benefit, indicating that using all candidate transformations is not necessary for effective calibration. To determine whether this gain simply comes from combining multiple transformations, we introduce Random-3, which randomly selects three transformations from the same candidate set while preserving the same calibration budget and subset partitioning as Entropy-3. This comparison isolates the effect of entropy-guided transformation selection from the benefit of using multiple transformations itself. Entropy-3 outperforms the average Random-3 result on both ID and corruption accuracy, showing that the gain cannot be explained by multi-transformation calibration alone. These results suggest that selecting a small subset of suitable transformations is more effective than either relying on a single transformation or combining transformations without a selection criterion. Detailed results across random transformation combinations are provided in Appendix G.

(a) Transformation selection strategy

Method	ID	C	Avg.
Baseline	64.50	31.17	47.84
Entropy-1	63.34	33.41	48.38
Entropy-3	**64.65**	**33.82**	**49.24**
Entropy-5	64.27	33.67	48.97
Entropy-9 (All)	64.26	33.16	48.71
Random-3	64.11	32.29	48.20

(b) Generalization across PTQ methods

Method	ID	C	Avg.	Method	ID	C	Avg.
APHQ-ViT	64.37	31.26	47.82	FIMA-Q	63.50	29.99	46.75
+MixUp	63.67	32.65	48.16	+MixUp	62.11	31.05	46.58
+AugMix	**64.41**	32.00	48.21	+AugMix	63.62	31.14	47.38
+HybridAug++	63.28	32.58	47.93	+HybridAug++	62.40	31.18	46.79
+PRIME	62.97	33.63	48.30	+PRIME	62.58	32.47	47.53
+**Entropy-K**	64.30	**34.23**	**49.27**	+**Entropy-K**	**63.66**	**32.77**	**48.22**

Table 2: **Ablation of transformation selection and generalization across PTQ methods under W3A3 quantization on ViT-S.** (a) Effect of the number and selection of calibration transformations. Entropy-3 achieves the best overall performance and outperforms Random-3. (b) Generalization of Entropy-K to APHQ-ViT and FIMA-Q against representative calibration transformations.

Generalization across PTQ methods. We examine whether the proposed transformation selection strategy generalizes beyond the LS-ViT baseline. Table 2(b) shows the results on APHQ-ViT and FIMA-Q. Across both PTQ methods, the effect of individual transformations varies considerably, and improvements in corruption robustness are often accompanied by degradation in ID performance. In contrast, Entropy-K consistently improves corruption robustness while maintaining ID performance close to the corresponding clean-calibration baselines. As a result, Entropy-K achieves the best average performance for both APHQ-ViT and FIMA-Q. These results suggest that Entropy-K is broadly applicable across different PTQ frameworks rather than being tailored to a particular quantization algorithm. Additional results are provided in Appendix H. Entropy-K also transfers beyond classification. On COCO with Mask R-CNN and Cascade Mask R-CNN (Swin-T/S backbones), it improves performance over an expanded pool of 13 candidates (Appendix I).

Feature-space preservation. We analyze how quantization affects the feature-space alignment between clean and corrupted inputs. Figure 6 visualizes the ImageNet and ImageNet-C feature distributions using UMAP (McInnes et al., 2018). The FP model exhibits substantial overlap between the two distributions, while the baseline quantized model shows a clear separation, indicating that clean calibration does not preserve the FP model's clean–corruption feature alignment. With Entropy-K, the two distributions overlap more substantially, yielding a feature distribution closer to that of the FP model. We further quantify this discrepancy using maximum mean discrepancy (MMD) (Gretton et al., 2012). The baseline quantized model yields an MMD of 0.1860, whereas Entropy-K reduces it to 0.1510, close to the FP model's 0.1444, indicating better preservation of the clean–corruption feature structure.

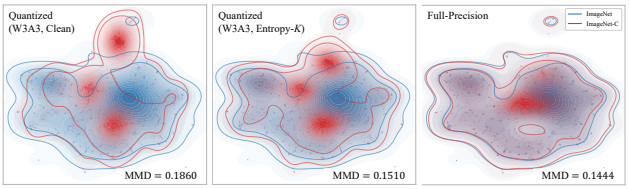


Figure 6: **Feature-space distributions of ImageNet and ImageNet-C.** Entropy-K reduces the clean–corruption discrepancy toward that of the FP model.

7 Conclusion

In this work, we studied corruption robustness in PTQ for ViTs from the perspective of calibration data. We showed that quantized models preserve substantially less of the FP model's performance under common corruptions than on clean inputs, and that different transformations of the calibration set can lead to markedly different robustness after quantization. Our analysis revealed that effective transformations tend to reduce activation energy, exhibiting a similar shift to that induced by corrupted inputs. We further associated this reduction with lower attention concentration and weaker directional alignment among value vectors in self-attention. Based on these observations, we introduced the *spatial entropy of difference energy*, S_{ent}, as a forward-free and model-independent indicator of transformation effectiveness. Using this score, we proposed *Entropy-K,* which selects multiple high-scoring transformations while preserving the original calibration budget and leaving the underlying PTQ algorithm unchanged. Across multiple ViT architectures, low-bit settings, and reconstruction-based PTQ methods, Entropy-K consistently improved corruption robustness while maintaining competitive in-distribution accuracy. These results highlight calibration-data transformation as a simple and broadly applicable direction for improving PTQ robustness without additional training data, retraining, or changes to the quantization algorithm.

REFERENCES

Hamidreza Yaghoubi Araghi, Parastoo Pilevar, and Ming C Lin. Recti-q: Feature-space rectification for out-of-distribution-robust quantized perception in edge robotics. *arXiv preprint arXiv:2607.18540*, 2026.

Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6154–6162. IEEE, 2018.

Ekin Dogus Cubuk, Barret Zoph, Jon Shlens, and Quoc Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Advances in Neural Information Processing Systems*, volume 33, pp. 18613–18624, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, 2009.

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. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, 13:723–773, 2012.

Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019.

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 *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020.

Hyunha Hwang, Xuan Truong Nguyen, and Hyuk-Jae Lee. Ls-vit: Least-squares hessian based block reconstruction for low-bit post-training quantization of vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 33588–33597, 2026.

Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2704–2713, 2018.

Yuhang Li, Ruihao Gong, Xu Tan, Yang Yang, Peng Hu, Qi Zhang, Fengwei Yu, Wei Wang, and Shi Gu. Brecq: Pushing the limit of post-training quantization by block reconstruction. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

Zhikai Li, Junrui Xiao, Lianwei Yang, and Qingyi Gu. Repq-vit: Scale reparameterization for post-training quantization of vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17181–17190. IEEE, 2023.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 21–37. Springer, 2016.

Yijiang Liu, Huanrui Yang, Zhen Dong, Kurt Keutzer, Li Du, and Shanghang Zhang. Noisyquant: Noisy bias-enhanced post-training activation quantization for vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20321–20330. IEEE, 2023.

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)*, pp. 10012–10022, October 2021.

Xiaofeng Mao, Gege Qi, Yuefeng Chen, Xiaodan Li, Ranjie Duan, Shaokai Ye, Yuan He, and Hui Xue. Towards robust vision transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12042–12051, 2022.

Leland McInnes, John Healy, Nathaniel Saul, and Lukas Grossberger. Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861, 2018.

Apostolos Modas, Rahul Rade, Guillermo Ortiz-Jiménez, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. Prime: A few primitives can boost robustness to common corruptions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

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)*, pp. 774–782, October 2021.

Markus Nagel, Rana Ali Amjad, Mart Van Baalen, Christos Louizos, and Tijmen Blankevoort. Up or down? adaptive rounding for post-training quantization. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2020.

Jaewoo Park, Jacky Chen Long Chai, Jaeho Yoon, and Andrew Beng Jin Teoh. Understanding the feature norm for out-of-distribution detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1557–1567, 2023.

Cuong Pham, Anh Dung Hoang, Cuong C Nguyen, Trung Le, Dinh Phung, Gustavo Carneiro, and Thanh-Toan Do. Metaaug: Meta-data augmentation for post-training quantization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 236–252, 2024.

Lei Qi, Peng Dong, Tan Xiong, Hui Xue, and Xin Geng. Doubleaug: Single-domain generalized object detector in urban via color perturbation and dual-style memory. *ACM Transactions on Multimedia Computing, Communications and Applications*, 20(5):1–20, 2024.

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 International Conference on Machine Learning (ICML)*, pp. 10347–10357. PMLR, 2021.

Xiuying Wei, Ruihao Gong, Yuhang Li, Xianglong Liu, and Fengwei Yu. Qdrop: Randomly dropping quantization for extremely low-bit post-training quantization. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022.

Zhuguanyu Wu, Jiaxin Chen, Hanwen Zhong, Di Huang, and Yunhong Wang. Adalog: Post-training quantization for vision transformers with adaptive logarithm quantizer. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 411–427. Springer, 2024.

Zhuguanyu Wu, Shihe Wang, Jiayi Zhang, Jiaxin Chen, and Yunhong Wang. Fima-q: Post-training quantization for vision transformers by fisher information matrix approximation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14891–14900, 2025a.

Zhuguanyu Wu, Jiayi Zhang, Jiaxin Chen, Jinyang Guo, Di Huang, and Yunhong Wang. Aphq-vit: Post-training quantization with average perturbation hessian based reconstruction for vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9686–9695, 2025b.

Yisong Xiao, Aishan Liu, Tianyuan Zhang, Haotong Qin, Jinyang Guo, and Xianglong Liu. Robustmq: benchmarking robustness of quantized models. *Visual Intelligence*, 1(1):30, 2023.

Xiaoran Xu, Jiangang Yang, Wenhui Shi, Siyuan Ding, Luqing Luo, and Jian Liu. Physaug: A physical-guided and frequency-based data augmentation for single-domain generalized object detection. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*, volume 39, pp. 21815–21823, 2025.

Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4084–4094, 2020.

Yeonguk Yu, Sungho Shin, Seongju Lee, Changhyun Jun, and Kyoobin Lee. Block selection method for using feature norm in out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15701–15711, 2023.

Zhihang Yuan, Chenhao Xue, Yiqi Chen, Qiang Wu, and Guangyu Sun. Ptq4vit: Post-training quantization for vision transformers with twin uniform quantization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 191–207. Springer, 2022.

Mehmet Kerim Yucel, Ramazan Gokberk Cinbis, and Pinar Duygulu. Hybridaugment++: Unified frequency spectra perturbations for model robustness. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Sangdoo 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6023–6032, 2019.

Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018.

Zhao Zhang, Yangcheng Gao, Jicong Fan, Zhongqiu Zhao, Yi Yang, and Shuicheng Yan. Selectq: Calibration data selection for post-training quantization. *Machine Intelligence Research*, 22(3): 499–510, 2025.

Yunshan Zhong, Jiawei Hu, Mingbao Lin, Mengzhao Chen, and Rongrong Ji. Is-vit: An inclusive stable method for post-training vits quantization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(2):1063–1080, 2026.

Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*, volume 34, pp. 13001–13008, 2020.

APPENDIX

LLM Usage Declaration Large language models (LLMs) were used solely for language refinement, and all research content was generated entirely by the authors.

A LIMITATIONS

Our forward-free proxy, S_{ent}, is not a complete predictor of transformation effectiveness. It captures the spatial distribution of transformation-induced differences while discarding other factors such as perturbation strength; we analyze this limitation in more detail in Appendix B. In addition, our correlation analyses are conducted over nine candidate transformations, selected to cover a diverse yet computationally tractable set of calibration perturbations. While this setting enables controlled comparison across heterogeneous transformations, it does not exhaust the broader space of possible transformations. Extending the analysis to a larger and more diverse candidate pool is left for future work.

B ADDITIONAL ANALYSIS OF THE FORWARD-FREE PROXY

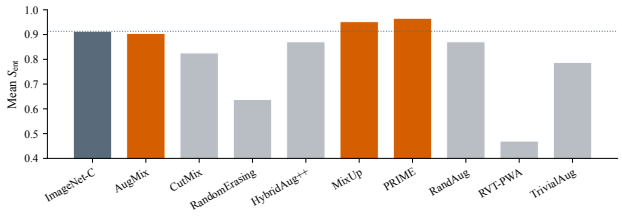


Figure 7: **Spatial entropy of ImageNet-C corruptions and candidate transformations.** The dotted line denotes the mean over 75 ImageNet-C corruption–severity conditions. Orange denotes the three transformations selected by Entropy-K.

Spatial entropy under common corruptions. We first examine whether common corruptions exhibit a spatial distribution of difference energy similar to that of the transformations selected by Entropy-K. As shown in Figure 7, the mean S_{ent} across 75 ImageNet-C corruption–severity conditions is 0.913. PRIME, MixUp, and AugMix, selected by Entropy-K, obtain scores of 0.966, 0.952, and 0.905, respectively, which are comparable to this value. In contrast, several lower-ranked transformations exhibit substantially lower spatial entropy. These results support the interpretation that Entropy-K favors transformations whose differences exhibit a spatial distribution similar to that induced by common corruptions.

Table 3: **Comparison of forward-free input-space proxies.** Spearman correlations are computed across nine transformations after averaging over ViT-S, DeiT-S, and Swin-S under W2A3. ρ_E and ρ_C denote correlations with the clean-relative middle-block activation-energy ratio and ImageNet-C accuracy gain, respectively.

Proxy	$\rho_E \downarrow$	$\rho_C \uparrow$
S_{ent} (16×16)	**-0.783**	**0.833**
S_{ent} (4×4)	-0.733	0.733
Normalized difference magnitude	0.150	0.367

Comparison with alternative forward-free proxies. We examine whether the effectiveness of S_{ent} can instead be explained by spatial granularity or the overall input perturbation magnitude. We compute the same spatial entropy using finer 4×4 regions to evaluate the effect of spatial granularity. We also consider the normalized difference magnitude

$$M(T) = \frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{x}\in\mathcal{D}} \frac{\|T(\boldsymbol{x}) - \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}, \tag{11}$$

which measures the overall amount of input change while discarding its spatial distribution. As shown in Table 3, S_{ent} with 16×16 regions exhibits the strongest architecture-mean correlation with both middle-block activation-energy reduction and ImageNet-C accuracy gain. The finer spatial entropy shows a similar but weaker trend, whereas normalized difference magnitude exhibits substantially weaker correlations. These results suggest that the spatial distribution of transformation-induced changes provides a useful signal that is not captured by their overall magnitude alone.

Table 4: **Correlation after controlling transformation magnitude.** Spearman correlation between S_{ent} and ImageNet-C accuracy gain across nine transformations on ViT-S under W2A3.

Setting	$M(T)$	$\rho_C \uparrow$
Standard setting	0.107–0.437	0.667
Magnitude matched	0.100	**0.883**

Controlling for transformation magnitude. Although magnitude alone provides a weak ranking signal across heterogeneous transformations, S_{ent} is invariant to uniform scaling of transformation-induced differences. For a scaled perturbation $\boldsymbol{\delta}' = \alpha\boldsymbol{\delta}$, where $\boldsymbol{\delta} = T(\boldsymbol{x}) - \boldsymbol{x}$, each regional difference energy is scaled as $e'_j = \alpha^2 e_j$. However, this factor cancels when computing the normalized spatial distribution:

$$p'_j = \frac{\alpha^2 e_j}{\sum_k \alpha^2 e_k} = p_j. \tag{12}$$

Consequently, S_{ent} captures the spatial distribution of transformation-induced differences while discarding their overall magnitudes. The observed entropy–robustness relationship could therefore still be influenced by differences in perturbation magnitude across transformations. To examine this, we explicitly match the input-space magnitude of all nine candidate transformations and repeat PTQ.

For each transformation T, we measure its normalized difference magnitude on a fixed reconstruction set $\mathcal{D}_{\text{rec}}$ using Eq. 11. We then rescale each transformation to obtain a magnitude-matched reconstruction set with $M(T) = 0.10$:

$$\widetilde{T}(\boldsymbol{x}) = \boldsymbol{x} + \beta_T \left(T(\boldsymbol{x}) - \boldsymbol{x}\right), \qquad \beta_T = \frac{0.10}{M(T)}. \tag{13}$$

The original magnitudes range from 0.107 to 0.437, so $\beta_T \leq 1$ for all transformations. We repeat ViT-S W2A3 quantization using the magnitude-matched transformations under the same protocol.

As shown in Table 4, matching perturbation magnitude does not remove the relationship between spatial entropy and corruption robustness. The Spearman correlation between S_{ent} and ImageNet-C accuracy gain increases from $\rho = 0.667$ in the original setting to $\rho = 0.883$ after magnitude matching. This result indicates that S_{ent} captures transformation properties relevant to corruption robustness beyond overall perturbation magnitude.

Table 5: **Gaussian-noise strength sweep on ViT-S under W2A3.** Increasing the noise strength changes perturbation magnitude while leaving spatial entropy nearly unchanged. ΔID and ΔC denote percentage-point changes from clean calibration.

Noise strength	$M(T)$	S_{ent}	ΔID	ΔC
Weak	0.033	0.9991	0.54	1.00
Medium	0.130	0.9986	-0.20	0.49
Strong	0.252	0.9982	-6.93	-1.28

A counterexample to spatial entropy alone. The magnitude-controlled comparison above shows that overall perturbation magnitude does not account for the robustness ranking across transformations. Instead, S_{ent} remains strongly associated with corruption robustness even after the magnitude differences are removed. However, high spatial entropy alone is not sufficient to guarantee an effective calibration transformation. A simple counterexample is additive Gaussian noise, whose perturbations are broadly distributed across the image and therefore yield $S_{\text{ent}} \approx 1$.

To this end, we use additive Gaussian noise at three strengths, with standard deviations $\sigma \in \{4/255, 16/255, 32/255\}$. We refer to these settings as *weak*, *medium*, and *strong*, respectively.

For each input, we sample pixel-wise noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2)$ and construct the transformed input as $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{\epsilon}$. Varying σ changes the perturbation magnitude while preserving the spatially distributed structure of Gaussian noise, resulting in nearly identical spatial entropy across the three settings.

As shown in Table 5, all three Gaussian-noise settings exhibit $S_{\text{ent}} \approx 1$, yet their quantization performance differs substantially. This result shows that high spatial entropy alone is not sufficient to characterize an effective calibration transformation. At the same time, the magnitude-controlled comparison above shows that the association between S_{ent} and corruption robustness remains strong even when overall perturbation magnitude is controlled across heterogeneous candidate transformations. Together, these results suggest that S_{ent} captures a useful spatial property that contributes to transformation effectiveness, while additional factors, such as perturbation strength, can also affect the resulting quantization behavior. Accordingly, we use S_{ent} as a simple forward-free ranking signal for calibration transformations rather than as a complete predictor of their effectiveness.

C Details of Calibration Transformations

To preserve the efficiency of PTQ, we exclude learnable augmentation methods that require an additional optimization procedure, such as MetaAug (Pham et al., 2024). In our PTQ setting, these transformations are used solely to construct transformed calibration inputs. Training-specific components, such as MixUp/CutMix label interpolation or auxiliary consistency losses, are omitted because PTQ reconstruction does not use classification labels. Example transformed images are shown in Figure 8. We briefly describe the nine transformations below, while their implementation details and parameter settings are summarized in Table 6.

Table 6: **Implementation details of image transformations.** Main operations and parameter settings for the nine transformations evaluated in our experiments.

Method	Main operation	Key setting
RandomErasing	Random rectangular masking	$p = 1$, scale $[0.02, 0.33]$
CutMix	Patch replacement between images	$\alpha = 1$
MixUp	Linear mixing of two images	$\alpha = 1$
RandAugment	Random geometric/appearance transforms	$N = 2$, $M = 9$
TrivialAugment	One random transform	31 magnitude bins
AugMix	Mixture of augmentation chains	sev. 3, width 3, depth 1–3
RVT-PWA	Patch-wise crop, noise, and flip	16×16, $p = 0.1$
HybridAugment++	LF/HF and amplitude–phase mixing	Balanced PS; level 3, $k = 3$, $\sigma_S = 0.5$, $\sigma_P = 1.0$
PRIME	Spectral, spatial, and color perturbations	width 3, depth 1–3

RandomErasing. RandomErasing (Zhong et al., 2020) randomly masks a rectangular region of an image. The transformation removes local visual content while leaving the remaining regions unchanged.

CutMix. CutMix (Yun et al., 2019) replaces a rectangular region of one image with a patch taken from another image. Unlike RandomErasing, which removes image content, CutMix fills the selected region with visual content from another sample.

MixUp. MixUp (Zhang et al., 2018) creates a new sample by linearly interpolating two images:

$$\tilde{\boldsymbol{x}} = \lambda \boldsymbol{x}_i + (1 - \lambda)\boldsymbol{x}_j, \tag{14}$$

where λ controls the contribution of each image. Unlike CutMix, which combines images within a local region, MixUp blends the two images over the entire spatial extent.

RandAugment. RandAugment (Cubuk et al., 2020) applies a small number of randomly selected transformations from a predefined augmentation space. The space includes geometric and appearance transformations such as rotation, translation, contrast adjustment, color adjustment, and posterization. A shared magnitude parameter controls the strength of the selected transformations.

TrivialAugment. TrivialAugment (Müller & Hutter, 2021) adopts a simpler augmentation strategy by randomly selecting a single transformation and its magnitude for each image. It avoids the policy search and additional hyperparameters required by more complex automated augmentation methods.

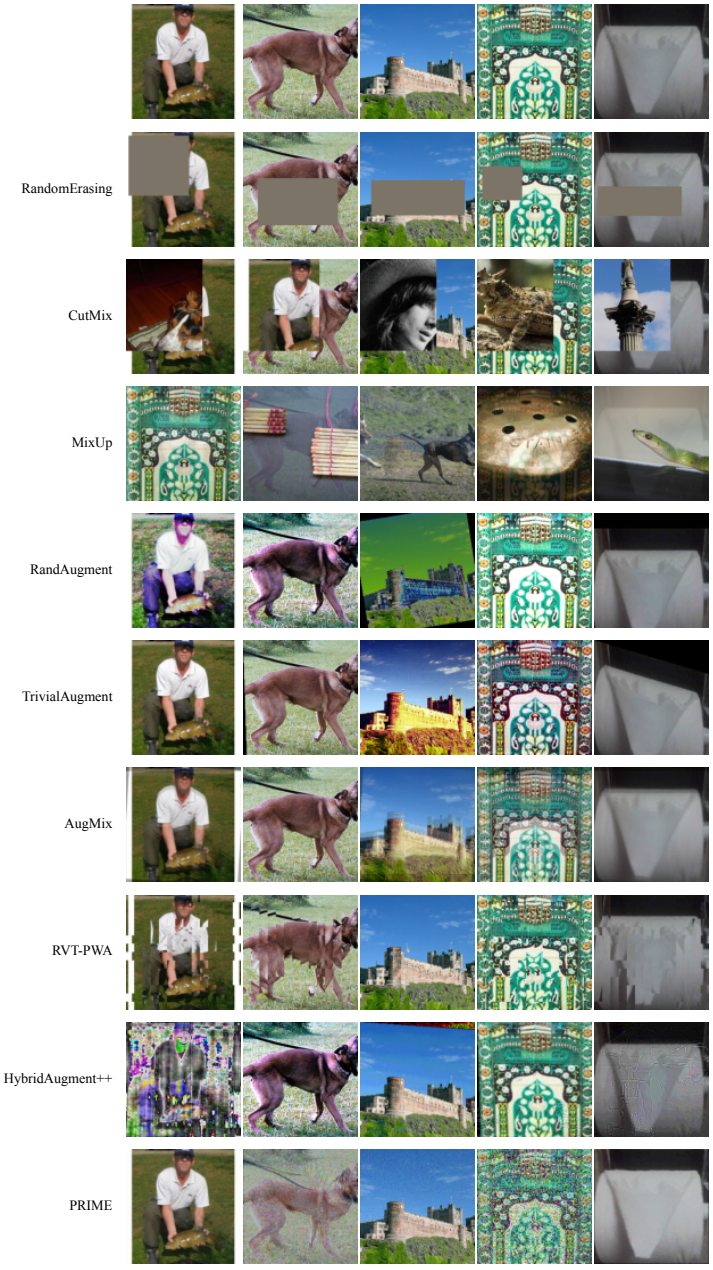


Figure 8: **Examples of calibration transformations.** Representative transformed images generated by the nine methods considered in our experiments.

AugMix. AugMix (Hendrycks et al., 2020) generates multiple randomly augmented variants of an image and combines them through stochastic mixing. The mixed augmentation is further combined with the original image, producing diverse perturbations while retaining substantial visual content from the original sample.

RVT-PWA. Patch-wise augmentation (PWA) from RVT (Mao et al., 2022) applies transformations independently to different image regions rather than applying a single transformation to the entire image. As a result, different patches can undergo different perturbations, introducing spatially heterogeneous variations within an image.

HybridAugment++. HybridAugment++ (Yucel et al., 2023) perturbs image frequency components by mixing low- and high-frequency information across views or samples. To preserve the

fixed calibration budget, we use a balanced PS configuration: half of the N samples use the single-image branch and the other half use the paired-image branch, keeping the total size at N instead of $2N$.

PRIME. PRIME (Modas et al., 2022) generates diverse perturbations using spectral, spatial, and color transformations. These transformations are randomly composed and mixed to produce a broad range of image variations while largely preserving the semantic content of the original image.

D Derivation of E_{attn} decomposition

We provide the derivation of the self-attention output energy decomposition in Eq. 5. For a single attention head, the output for the i-th query token is $\boldsymbol{y}_i = \sum_{j=1}^{K} a_{ij}\boldsymbol{v}_j$. The attention-output energy can be expanded as

$$
\begin{aligned}
E_{\text{attn}} &= \|AV\|_F^2 \\
&= \sum_{i=1}^{Q} \left\| \sum_{j=1}^{K} a_{ij}\boldsymbol{v}_j \right\|_2^2 \\
&= \sum_{i=1}^{Q} \left\langle \sum_{j=1}^{K} a_{ij}\boldsymbol{v}_j, \sum_{k=1}^{K} a_{ik}\boldsymbol{v}_k \right\rangle \\
&= \sum_{i=1}^{Q} \left(\sum_{j=1}^{K}\sum_{k=1}^{K} a_{ij}a_{ik}\langle \boldsymbol{v}_j, \boldsymbol{v}_k \rangle \right) \\
&= \sum_{i=1}^{Q} \left(\sum_{\substack{j,k=1 \\ j=k}}^{K} a_{ij}a_{ik}\langle \boldsymbol{v}_j, \boldsymbol{v}_k \rangle + \sum_{\substack{j,k=1 \\ j\neq k}}^{K} a_{ij}a_{ik}\langle \boldsymbol{v}_j, \boldsymbol{v}_k \rangle \right) \\
&= \underbrace{\sum_{i=1}^{Q}\sum_{j=1}^{K} a_{ij}^2 \|\boldsymbol{v}_j\|_2^2}_{D} + \underbrace{\sum_{i=1}^{Q}\sum_{\substack{j,k=1 \\ j\neq k}}^{K} a_{ij}a_{ik}\langle \boldsymbol{v}_j, \boldsymbol{v}_k \rangle}_{X}.
\end{aligned}
\tag{15}
$$

The term D collects the $j = k$ terms, corresponding to the diagonal entries of the value-vector Gram matrix $VV^\top$. The term X collects the $j \neq k$ terms, corresponding to its off-diagonal entries and capturing interactions between different value vectors.

E Direct Activation Scaling

To isolate the effect of activation magnitude from other transformation-induced changes, we directly scale the outputs of selected Transformer blocks without applying image transformations. For a block output h_b, we apply

$$h_b' = (1 - r)h_b, \tag{16}$$

where r denotes the reduction ratio. The scaled activation is propagated to subsequent blocks, while the activation quantization parameters are calibrated beforehand using clean inputs and remain fixed throughout the experiment. Table 7 compares scaling all Transformer blocks with scaling only the middle blocks (B4–B7) on ViT-S. Scaling all blocks progressively degrades both ImageNet and ImageNet-C accuracy as the reduction ratio increases. Scaling only the middle blocks is less disruptive, but reductions of 1–5% do not consistently improve corruption accuracy, while a 10% reduction clearly degrades performance. These results show that reducing activation magnitude alone is not sufficient to improve corruption robustness. Thus, the lower middle-block activation energy observed for effective calibration transformations is better viewed as an associated characteristic rather than a sufficient mechanism for their robustness gains.

Route	Reduction	W2A3		W2A4		W3A3	
		ID	C	ID	C	ID	C
All	0%	42.02	15.42	59.59	27.58	64.50	31.17
	1%	41.83	15.08	58.78	27.10	64.39	30.91
	3%	40.11	14.12	56.79	25.60	63.23	29.63
	5%	36.64	12.48	52.92	22.89	59.58	27.18
	10%	19.62	5.30	32.68	11.09	38.16	14.00
Middle	0%	42.02	15.42	59.59	27.58	64.50	31.17
	1%	42.07	15.35	59.70	27.14	64.48	30.93
	3%	41.89	15.35	59.33	27.20	64.40	30.78
	5%	41.51	14.91	58.50	26.88	64.00	30.62
	10%	39.10	13.76	56.06	24.62	61.86	28.74

Table 7: ImageNet and ImageNet-C accuracy (%) of ViT-S under direct activation-magnitude scaling during reconstruction and evaluation. *All* scales the outputs of all blocks, whereas *Middle* scales only B4–B7.

F Category-wise Corruption Robustness and Noise Sensitivity

Method	W/A	ViT-S				DeiT-S				Swin-S			
		Noise	Blur	Weather	Digital	Noise	Blur	Weather	Digital	Noise	Blur	Weather	Digital
FP	32/32	57.27	50.82	61.06	62.67	56.06	45.63	62.57	60.88	63.96	53.78	68.42	67.35
Baseline	3/3	25.31	27.71	34.30	35.89	32.32	31.71	44.12	42.85	38.59	37.77	52.60	48.51
Entropy-K	3/3	30.38	29.02	36.75	38.27	38.39	33.89	45.73	45.12	42.18	38.43	54.92	50.72
Baseline	2/4	21.13	24.17	31.22	32.18	30.17	28.92	40.87	39.21	36.60	34.46	49.10	44.92
Entropy-K	2/4	27.79	25.06	32.93	33.45	35.75	30.51	43.01	40.18	38.76	34.15	50.54	46.75
Baseline	2/3	8.71	13.09	18.44	19.75	17.60	18.42	27.21	26.92	24.64	26.25	39.07	34.42
Entropy-K	2/3	14.71	14.93	19.86	20.83	23.44	21.77	31.09	30.72	28.71	27.48	41.70	37.36

Table 8: **Category-wise corruption accuracy (%).** ImageNet-C accuracy across four corruption categories for ViT-S, DeiT-S, and Swin-S under different quantization settings.

Table 8 reports ImageNet-C accuracy across four corruption categories. Although the absolute difficulty of each category varies for the FP models, noise consistently exhibits the largest quantization-induced degradation across all three architectures and bit-width settings. This result suggests that noise corruptions are particularly sensitive to quantization.

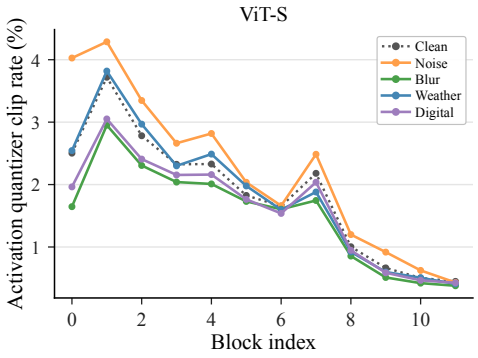


Figure 9: **Activation clipping across network depth.** Clipping rates of activation quantizers for clean inputs and four ImageNet-C corruption categories under W2A3 quantization.

To better understand the pronounced degradation under noise, we examine activation clipping across network depth. Figure 9 shows the fraction of activation values that exceed the calibrated quantization range. For ViT-S, noise produces a clear increase in clipping in the early blocks, whereas the

gap becomes substantially smaller in later blocks. The other corruption categories remain closer to clean inputs or exhibit lower clipping rates in the early blocks. This pattern suggests that the vulnerability to noise is associated with an early mismatch between corrupted activation distributions and the quantization range determined from clean calibration data. We further examine whether Entropy-K alleviates this mismatch. As shown in Table 9, Entropy-K reduces the average clipping rate under noise across all three architectures, from 2.208% to 2.098% on ViT-S, from 1.689% to 1.580% on DeiT-S, and from 2.217% to 1.922% on Swin-S. These results show that the calibration transformations selected by Entropy-K reduce activation clipping under noise, which exhibits the largest quantization-induced degradation.

Method	ViT-S		DeiT-S		Swin-S	
	Clean	Noise	Clean	Noise	Clean	Noise
Baseline	1.828	2.208	1.437	1.689	1.980	2.217
Entropy-K	1.834	2.098	1.243	1.580	1.727	1.922

Table 9: **Activation clipping rates (%) on clean and noise inputs under W2A3 quantization.** Entropy-K reduces noise-induced activation clipping across all three architectures.

G Additional Analysis of Transformation Selection

Method	Trial	ViT-S		DeiT-S	
		ID	C	ID	C
Random-3	1	63.75	33.41	70.08	40.44
	2	63.60	31.53	70.28	40.59
	3	64.09	32.69	69.86	40.24
	4	64.58	33.31	69.64	40.23
	5	64.33	31.34	69.88	39.82
	6	64.01	32.58	70.05	41.00
	7	63.86	31.60	69.79	39.75
	8	64.16	33.07	69.84	40.39
	9	64.35	31.56	69.96	39.58
	10	64.33	31.79	69.31	39.02
	Avg.	64.11	32.29	69.87	40.11
Entropy-3		**64.65**	**33.82**	**69.93**	**40.94**

(a) W3A3

Method	Trial	ViT-S		DeiT-S	
		ID	C	ID	C
Random-3	1	42.77	16.96	56.05	27.41
	2	41.84	17.36	55.08	27.37
	3	41.86	17.08	54.03	26.21
	4	43.16	15.73	54.77	25.14
	5	41.97	16.24	55.74	26.75
	6	43.20	16.22	54.80	25.10
	7	41.60	15.76	53.96	25.42
	8	41.76	16.92	55.16	26.60
	9	41.41	15.70	53.33	24.14
	10	41.59	16.99	54.57	26.44
	Avg.	42.12	16.50	54.75	26.06
Entropy-3		**42.47**	**17.78**	**54.76**	**26.98**

(b) W2A3

Table 10: **Comparison between random transformation selection and Entropy-K.** Random-3 is evaluated over ten independent trials under W3A3 and W2A3 on ViT-S and DeiT-S.

In the main paper, we compare Entropy-3 with Random-3 under W3A3 quantization on ViT-S to examine whether the gain arises from selecting suitable transformations rather than simply combining multiple transformations. Here, we provide a more detailed comparison by repeating Random-3 over ten independent trials and extending the evaluation to DeiT-S and the W2A3 setting.

Table 10 reports the results of all random trials together with their mean. Under W3A3, Entropy-3 improves ImageNet-C accuracy over the Random-3 mean by 1.53 percentage points on ViT-S and 0.83 points on DeiT-S. At the same time, ImageNet accuracy is maintained or slightly improved, with gains of 0.54 and 0.06 points, respectively. A similar trend is observed under W2A3, where Entropy-3 improves ImageNet-C accuracy by 1.28 points on ViT-S and 0.92 points on DeiT-S, while the corresponding ID gains are 0.35 and 0.01 points. These results show that the advantage of Entropy-3 is not specific to the single Random-3 result reported in the main paper. Across both architectures and bit-width settings, Entropy-3 consistently exceeds the Random-3 mean in corruption accuracy while maintaining comparable or better ID accuracy. This comparison further supports the conclusion that the improvement arises from entropy-guided transformation selection rather than from multi-transformation calibration alone.

H ADDITIONAL GENERALIZATION PERFORMANCE ACROSS PTQ METHODS

Method	W/A	ViT-S		DeiT-S		Swin-S		Avg.
		ID	C	ID	C	ID	C	
Full-Prec	32/32	81.38	58.00	79.86	56.30	83.32	63.34	70.37
APHQ-ViT	3/3	64.37	31.26	68.38	37.28	77.35	44.29	53.82
+RandomErasing	3/3	64.10	31.73	68.81	37.76	77.28	44.12	53.97
+CutMix	3/3	63.82	30.15	70.40	39.78	78.35	45.95	54.74
+MixUp	3/3	63.67	32.65	69.41	39.97	77.97	47.15	55.14
+RandAugment	3/3	64.36	32.33	68.74	38.08	77.46	44.31	54.21
+TrivialAugment	3/3	64.32	32.24	68.74	38.44	77.66	46.23	54.60
+AugMix	3/3	64.41	32.00	69.07	38.91	77.93	45.77	54.68
+RVT-PWA	3/3	64.39	31.15	68.46	37.45	77.47	44.06	53.83
+HybridAugment++	3/3	63.28	32.58	69.46	40.62	78.05	46.45	55.07
+PRIME	3/3	62.97	33.63	69.17	**40.79**	77.52	46.38	55.08
+Entropy-K	3/3	64.30	**34.23**	69.12	40.12	77.72	**47.40**	**55.48**
APHQ-ViT	2/4	55.95	24.46	63.90	33.25	72.92	39.11	48.27
+RandomErasing	2/4	56.49	24.87	64.11	33.60	73.33	39.89	48.72
+CutMix	2/4	56.03	23.59	65.50	34.84	74.72	40.63	49.22
+MixUp	2/4	54.89	25.41	64.37	35.68	73.96	41.69	49.33
+RandAugment	2/4	56.28	25.67	63.77	33.78	73.36	40.12	48.83
+TrivialAugment	2/4	55.56	25.47	64.06	34.06	73.51	40.35	48.84
+AugMix	2/4	56.55	25.81	64.40	35.28	73.54	41.36	49.49
+RVT-PWA	2/4	56.19	24.23	63.78	33.56	73.28	39.48	48.42
+HybridAugment++	2/4	53.46	24.83	64.25	35.54	73.79	41.68	48.93
+PRIME	2/4	53.22	26.18	63.34	**35.99**	72.93	40.66	48.72
+Entropy-K	2/4	55.56	**26.96**	64.10	35.67	73.45	**42.06**	**49.63**
APHQ-ViT	2/3	43.07	15.58	50.33	20.68	65.95	29.89	37.58
+RandomErasing	2/3	41.97	15.36	51.92	21.93	65.68	30.26	37.85
+CutMix	2/3	42.49	14.97	55.56	24.81	68.52	31.95	**39.72**
+MixUp	2/3	41.57	16.35	53.76	25.40	66.76	32.21	39.34
+RandAugment	2/3	43.05	16.94	51.49	22.06	65.98	30.24	38.29
+TrivialAugment	2/3	42.51	16.39	51.76	22.55	66.63	31.06	38.48
+AugMix	2/3	42.92	16.62	52.89	24.24	67.05	32.24	39.33
+RVT-PWA	2/3	42.26	15.65	50.59	21.14	65.77	29.78	37.53
+HybridAugment++	2/3	39.39	15.85	53.15	25.71	66.71	32.44	38.88
+PRIME	2/3	39.10	16.76	52.58	**26.36**	65.18	32.01	38.66
+Entropy-K	2/3	42.03	**17.46**	53.28	25.57	66.64	**32.96**	39.66

(a) APHQ-ViT

Method	W/A	ViT-S		DeiT-S		Swin-S		Avg.
		ID	C	ID	C	ID	C	
Full-Prec	32/32	81.38	58.00	79.86	56.30	83.32	63.34	70.37
FIMA-Q	3/3	63.50	29.99	68.68	37.56	77.27	43.27	53.38
+RandomErasing	3/3	63.26	29.58	69.05	37.83	77.14	43.28	53.36
+CutMix	3/3	63.02	29.28	70.58	39.99	78.28	44.90	54.34
+MixUp	3/3	62.11	31.05	69.59	40.23	77.70	**46.45**	54.52
+RandAugment	3/3	63.38	30.82	68.90	38.69	77.28	44.01	53.85
+TrivialAugment	3/3	63.30	30.89	69.14	38.64	77.56	44.38	53.99
+AugMix	3/3	63.62	31.14	69.49	39.54	77.57	45.00	54.39
+RVT-PWA	3/3	63.33	30.21	68.90	37.72	77.30	43.35	53.47
+HybridAugment++	3/3	62.40	31.18	69.71	41.02	77.61	45.77	54.62
+PRIME	3/3	62.58	32.47	69.68	**41.32**	77.50	45.89	54.91
+Entropy-K	3/3	63.66	**32.77**	69.71	40.87	77.82	46.14	**55.16**
FIMA-Q	2/4	57.92	26.19	65.13	34.93	73.26	39.72	49.52
+RandomErasing	2/4	57.83	25.45	65.10	35.18	72.60	38.75	49.15
+CutMix	2/4	58.27	24.12	66.95	36.91	74.66	40.68	50.26
+MixUp	2/4	57.37	27.49	66.43	36.83	74.02	**41.76**	50.65
+RandAugment	2/4	57.69	27.10	65.26	35.38	73.02	39.35	49.63
+TrivialAugment	2/4	57.66	26.75	65.01	34.97	73.27	40.31	49.66
+AugMix	2/4	57.82	26.93	65.72	36.39	73.75	41.25	50.31
+RVT-PWA	2/4	58.19	26.07	65.25	35.20	73.35	39.59	49.61
+HybridAugment++	2/4	56.05	27.07	65.48	36.86	73.57	41.27	50.05
+PRIME	2/4	55.41	28.20	64.96	36.81	73.30	40.08	49.79
+Entropy-K	2/4	58.00	**28.90**	65.76	**37.00**	73.43	40.99	**50.68**
FIMA-Q	2/3	40.72	14.82	51.21	22.37	64.88	28.33	37.05
+RandomErasing	2/3	40.50	14.15	52.60	22.83	64.02	26.55	36.77
+CutMix	2/3	40.84	14.25	56.36	25.88	67.50	29.65	**39.08**
+MixUp	2/3	39.29	15.08	54.57	26.54	66.34	31.18	38.83
+RandAugment	2/3	40.21	15.50	52.38	23.49	64.42	28.30	37.38
+TrivialAugment	2/3	39.48	14.56	51.54	23.35	65.08	29.32	37.22
+AugMix	2/3	41.28	15.73	53.57	24.78	65.95	30.62	38.66
+RVT-PWA	2/3	41.11	14.80	51.81	22.25	64.40	28.21	37.10
+HybridAugment++	2/3	38.75	16.00	54.33	**26.78**	65.72	31.41	38.83
+PRIME	2/3	38.63	16.30	52.94	26.76	64.81	29.99	38.24
+Entropy-K	2/3	40.82	**16.63**	53.05	26.14	65.82	**31.81**	39.04

(b) FIMA-Q

Table 11: **Results across different PTQ methods.** ImageNet and ImageNet-C accuracy of APHQ-ViT and FIMA-Q with different calibration transformations on ViT-S, DeiT-S, and Swin-S under W3A3, W2A4, and W2A3. Avg. denotes the mean of the six ImageNet and ImageNet-C accuracy values across the three architectures.

Table 11 presents the results of applying calibration transformations to APHQ-ViT and FIMA-Q across three architectures and three bit-width settings. Across both PTQ methods, the effect of individual transformations varies considerably with architecture and quantization precision. A transformation that improves corruption accuracy in one setting does not necessarily provide the same benefit in another, and larger robustness gains are sometimes accompanied by noticeable ID degradation. This variation is consistent with the setting-dependent effectiveness of individual calibration transformations observed in the main paper.

In contrast, Entropy-K consistently improves ImageNet-C accuracy over the corresponding clean-calibration baseline across all evaluated architectures under W3A3, W2A4, and W2A3. This improvement is also reflected in the overall Avg. score: compared with the corresponding baselines, Entropy-K improves Avg. by 1.66, 1.36, and 2.08 points for APHQ-ViT, and by 1.78, 1.16, and 1.99 points for FIMA-Q under W3A3, W2A4, and W2A3, respectively. At the same time, Entropy-K maintains ID performance close to the corresponding clean-calibration baseline in most settings. This contrasts with several individual transformations that achieve larger corruption gains at the cost of substantial ID degradation. Moreover, the effectiveness of individual transformations varies noticeably across PTQ methods and bit-widths, whereas Entropy-K exhibits a more stable improvement across these settings. This consistency indicates that the proposed selection strategy is less sensitive to the specific reconstruction method or quantization precision. As a result, Entropy-K achieves either the best or second-best average performance across all six PTQ-method and bit-width combinations reported in the table.

These results show that the proposed transformation selection strategy generalizes across different reconstruction-based PTQ methods. Despite the different reconstruction objectives of APHQ-ViT and FIMA-Q, Entropy-K consistently improves corruption robustness for both methods. This supports transformation selection as a complementary component that can be applied on top of different reconstruction-based PTQ frameworks.

I Quantization Results on Downstream Tasks

We evaluate whether the proposed calibration strategy generalizes to downstream vision tasks using Mask R-CNN (He et al., 2017) and Cascade Mask R-CNN (Cai & Vasconcelos, 2018) with Swin-T and Swin-S backbones. For this setting, we expand the candidate pool with transformations used for object detection and semantic segmentation, including DoubleAUG-CP (Qi et al., 2024), SSD-Photometric (Liu et al., 2016), PhysAug (Xu et al., 2025), and FDA (Yang & Soatto, 2020). Together with the nine transformations considered for ImageNet classification, this yields 13 candidate transformations in total. We recompute S_{ent} over this expanded candidate set rather than reusing the transformations selected for ImageNet classification. With $K = 3$, PRIME, FDA, and MixUp obtain the highest scores and are selected following the same fixed-budget Entropy-K procedure.

Method	W/A	Mask R-CNN								Cascade Mask R-CNN							
		Swin-T				Swin-S				Swin-T				Swin-S			
		ID		C		ID		C		ID		C		ID		C	
		AP^b	AP^m	AP^b	AP^m	AP^b	AP^m	AP^b	AP^m	AP^b	AP^m	AP^b	AP^m	AP^b	AP^m	AP^b	AP^m
Full-Prec	32/32	46.00	41.60	15.40	13.49	48.50	43.20	17.85	15.54	50.40	43.70	17.28	14.56	52.00	45.00	18.89	15.85
Baseline	3/3	8.00	9.30	1.27	1.45	20.20	22.50	4.83	5.19	37.60	33.80	7.89	6.91	38.90	34.90	9.43	8.37
Entropy-K	3/3	**15.90**	**17.80**	**3.33**	**3.60**	**24.20**	**27.50**	**6.73**	**7.32**	**40.40**	**36.20**	**9.99**	**8.71**	**43.40**	**38.80**	**12.69**	**10.98**

Table 12: **Quantization results (%) on COCO object detection and instance segmentation.** ID denotes clean performance, and C denotes the mean over 15 corruption types at severity 5. AP^b and AP^m denote bounding-box AP and mask AP, respectively.

Table 12 reports bounding-box and mask AP on clean COCO inputs and corrupted inputs at severity 5. Entropy-K improves both AP^b and AP^m over the clean-calibration baseline across all evaluated detectors, backbones, and evaluation conditions. For example, with Mask R-CNN and Swin-T, Entropy-K improves clean AP^b and AP^m by $+7.9$ and $+8.5$ points, respectively, while improving the corresponding corrupted performance by $+2.06$ and $+2.15$ points. Similar improvements are observed with Cascade Mask R-CNN, supporting the applicability of entropy-guided calibration beyond image classification.

J Calibration Budget Analysis

Budget	Method	ViT-S		DeiT-S		Swin-S		Avg.
		ID	C	ID	C	ID	C	
1,024	Baseline	64.50	31.17	69.18	38.11	77.91	44.76	54.27
	Entropy-K	64.65	33.82	69.93	40.94	78.24	46.85	**55.74**
2,048	Baseline	66.43	33.07	70.34	39.83	78.87	45.99	55.76
	Entropy-K	66.39	36.12	71.02	42.36	78.96	47.75	**57.10**
4,096	Baseline	67.75	34.82	71.44	40.85	79.53	46.78	56.86
	Entropy-K	67.46	37.49	71.64	43.32	79.32	48.63	**57.98**

Table 13: **Effect of calibration set size.** ImageNet ID and ImageNet-C accuracy of the baseline and Entropy-K when varying the number of calibration images used for reconstruction from 1,024 to 4,096. Avg. denotes the mean of the six ID and C accuracy values across the three architectures.

We further examine whether the benefit of Entropy-K depends on the calibration budget. Table 13 compares the clean-calibration baseline and Entropy-K using 1,024, 2,048, and 4,096 calibration images. Increasing the calibration set size improves both ID and corruption accuracy for the baseline, while Entropy-K provides consistent additional gains across all three budgets with comparable ID performance. Compared with the corresponding baseline, the overall Avg. improves by $+1.47$, $+1.34$, and $+1.12$ percentage points with 1,024, 2,048, and 4,096 calibration images, respectively. Notably, Entropy-K with 2,048 calibration images achieves higher ImageNet-C accuracy than the baseline with 4,096 images across all three architectures. This indicates that transformation selection provides benefits beyond simply increasing the calibration-set size. Overall, calibration-set composition remains important even as the calibration budget increases.

K Additional Quantitative Results

Method	W/A	ViT-S		ViT-B		DeiT-T		DeiT-S		DeiT-B		Swin-S		Swin-B		Avg.
		ID	C	ID	C	ID	C	ID	C	ID	C	ID	C	ID	C	
Full-Prec	32/32	81.38	58.00	85.10	69.22	72.17	43.76	79.86	56.30	81.98	61.15	83.32	63.34	85.27	67.02	70.56
PTQ4ViT	3/3	0.85	0.42	0.25	0.18	2.69	1.02	0.27	0.20	30.28	12.16	39.68	12.15	38.57	12.12	10.77
RepQ-ViT	3/3	0.48	0.25	0.25	0.18	1.03	0.51	4.70	1.38	7.45	1.61	4.78	0.67	1.29	0.51	1.79
NoisyQuant	3/3	0.39	0.22	0.33	0.17	1.50	0.51	1.65	0.54	9.82	2.97	39.69	8.48	15.49	3.10	6.06
AdaLog	3/3	13.92	3.55	38.02	12.79	31.42	10.43	24.33	7.16	57.57	25.66	64.51	29.93	69.58	35.20	30.29
I&S-ViT	3/3	44.81	16.62	62.79	30.09	40.85	13.70	56.80	24.26	72.83	41.62	73.90	36.25	69.34	33.27	44.08
APHQ-ViT	3/3	64.37	31.26	73.67	41.81	56.08	25.78	68.38	37.28	76.15	49.01	77.35	44.29	77.95	50.02	55.24
FIMA-Q	3/3	63.50	29.99	74.54	43.69	55.81	25.72	68.68	37.56	76.27	48.83	77.27	43.27	78.67	48.72	55.18
LS-ViT	3/3	64.50	31.17	74.92	45.10	55.86	25.41	69.18	38.11	76.48	49.20	77.91	44.76	79.50	49.85	55.85
+RandomErasing	3/3	64.07	30.91	74.82	44.76	55.33	25.34	69.07	38.25	76.08	48.02	77.55	43.66	79.10	49.37	55.45
+CutMix	3/3	63.88	30.98	75.30	44.61	57.09	26.56	70.73	39.73	77.00	48.87	78.80	45.24	80.04	51.03	56.42
+MixUp	3/3	63.55	32.46	74.39	45.99	55.64	27.24	69.75	40.17	76.25	50.59	78.40	46.78	78.97	50.76	56.50
+AugMix	3/3	64.41	32.49	74.85	46.48	56.15	26.66	69.76	39.74	76.42	50.14	78.29	45.94	79.31	50.73	56.53
+RandAugment	3/3	64.54	31.94	74.91	45.43	55.70	25.59	69.24	38.91	76.33	49.60	77.64	44.38	79.14	50.48	55.99
+TrivialAugment	3/3	64.58	32.04	74.94	45.81	55.81	26.18	69.32	38.84	76.43	49.42	78.01	45.32	79.46	51.03	56.23
+HybridAugment++	3/3	63.47	32.27	74.53	46.72	56.21	26.91	70.13	41.08	76.64	50.59	78.13	46.57	79.37	51.33	56.71
+RVT-PWA	3/3	64.55	31.28	74.96	44.91	56.00	25.75	68.89	38.02	76.47	49.08	77.89	44.66	79.38	49.83	55.83
+PRIME	3/3	63.34	33.41	74.67	48.25	55.71	27.74	69.54	41.41	76.30	51.11	77.72	46.28	79.15	50.80	56.82
+**Entropy-K**	3/3	64.65	33.82	74.92	48.49	56.32	27.71	69.93	40.94	76.75	50.84	78.24	46.85	79.30	51.15	**57.14**
APHQ-ViT	2/4	55.95	24.46	68.31	36.33	49.71	20.68	63.90	33.25	72.61	43.98	72.92	39.11	72.67	43.45	49.81
FIMA-Q	2/4	57.92	26.19	70.15	39.83	51.39	22.45	65.13	34.93	73.37	45.43	73.26	39.72	72.76	40.57	50.94
LS-ViT	2/4	59.59	27.58	70.78	41.26	51.65	22.74	65.22	35.10	73.68	45.87	74.75	41.58	74.94	44.58	52.09
+RandomErasing	2/4	58.71	26.66	70.68	41.12	50.92	21.91	65.45	35.24	73.15	45.08	73.98	40.84	75.14	44.68	51.68
+CutMix	2/4	59.10	26.38	71.40	41.68	52.19	22.44	67.19	37.02	74.56	46.12	76.00	42.67	76.22	45.36	52.74
+MixUp	2/4	58.75	28.51	70.97	43.34	51.36	23.83	66.32	37.06	73.95	48.35	75.34	43.53	74.54	43.96	52.84
+RandAugment	2/4	59.27	28.06	71.01	42.23	51.30	22.45	65.53	35.61	73.73	46.15	74.72	41.86	74.96	45.70	52.33
+TrivialAugment	2/4	58.95	27.87	70.17	41.98	51.22	22.34	65.37	35.79	73.47	46.54	74.93	42.16	75.10	45.69	52.26
+AugMix	2/4	59.67	28.28	71.13	43.00	51.77	23.51	66.20	36.70	73.76	47.39	74.82	42.78	75.07	45.89	52.86
+RVT-PWA	2/4	59.64	27.42	70.73	41.09	51.81	22.50	65.84	35.24	73.67	45.94	74.44	41.89	75.20	44.97	52.17
+HybridAugment++	2/4	57.09	27.53	70.13	42.46	50.64	23.10	65.98	37.44	73.70	47.37	74.84	43.16	75.14	46.38	52.50
+PRIME	2/4	57.21	29.48	70.11	43.60	50.35	24.54	65.57	37.06	73.48	47.78	74.55	42.02	74.36	44.66	52.48
+**Entropy-K**	2/4	59.04	29.94	70.88	44.28	51.83	24.34	66.26	37.47	73.85	48.08	74.95	42.80	75.02	44.86	**53.11**
APHQ-ViT	2/3	43.07	15.58	53.12	20.22	37.83	13.40	50.33	20.68	66.19	34.31	65.95	29.89	64.44	31.72	39.05
FIMA-Q	2/3	40.72	14.82	55.61	22.66	37.90	13.97	51.21	22.37	66.46	34.96	64.88	28.33	63.70	28.99	39.04
LS-ViT	2/3	42.02	15.42	56.76	23.43	38.63	14.26	52.45	22.87	67.00	35.73	67.73	31.37	67.76	34.16	40.69
+RandomErasing	2/3	42.54	15.43	56.47	23.61	36.33	12.80	53.37	23.16	66.44	33.63	67.38	30.73	66.49	32.05	40.03
+CutMix	2/3	43.04	15.26	57.78	23.82	39.73	14.34	56.84	26.15	68.16	35.05	69.72	32.78	69.69	34.05	41.89
+MixUp	2/3	41.95	16.44	56.73	26.26	37.91	15.85	55.19	26.84	66.66	38.10	68.77	33.73	65.89	32.32	41.62
+RandAugment	2/3	41.18	15.76	57.37	24.65	37.98	14.43	53.18	24.16	67.01	36.37	67.34	31.73	68.20	34.45	40.99
+TrivialAugment	2/3	42.06	15.70	55.94	24.10	37.64	14.11	53.53	24.30	67.14	36.43	68.52	32.63	67.68	34.52	41.02
+AugMix	2/3	43.73	16.85	57.17	25.61	38.77	14.91	54.35	25.18	67.30	37.49	68.90	33.72	68.67	34.75	41.96
+RVT-PWA	2/3	43.26	15.88	57.28	24.17	38.20	14.07	52.99	22.95	66.94	35.98	67.39	31.96	68.24	33.83	40.94
+HybridAugment++	2/3	40.40	16.38	55.87	25.18	38.23	14.95	55.61	27.47	67.11	38.58	68.99	34.24	67.81	35.76	41.90
+PRIME	2/3	40.92	17.87	56.65	27.68	37.38	16.07	53.89	27.31	66.83	38.65	67.26	32.75	65.56	33.47	41.59
+**Entropy-K**	2/3	42.47	17.78	57.32	27.83	38.40	16.07	54.76	26.98	67.25	39.24	68.36	34.15	67.72	34.52	**42.35**

Table 14: **Full experimental results on ImageNet and ImageNet-C.** ID and C denote ImageNet and ImageNet-C accuracy, respectively. Compared with the main table, we additionally report W2A4 results and include PTQ4ViT, RepQ-ViT, NoisyQuant, AdaLog, and I&S-ViT under W3A3.

Table 14 provides the full quantitative results corresponding to the main comparison. The table extends the main results in two directions. First, we additionally report W2A4 to complement the W3A3 and W2A3 settings evaluated in the main text. Second, under W3A3, we include additional ViT-specific PTQ baselines, including PTQ4ViT (Yuan et al., 2022), RepQ-ViT (Li et al., 2023), NoisyQuant (Liu et al., 2023), AdaLog (Wu et al., 2024), and I&S-ViT (Zhong et al., 2026), in addition to APHQ-ViT (Wu et al., 2025b), FIMA-Q (Wu et al., 2025a), and LS-ViT (Hwang et al., 2026). All methods are evaluated on the same seven architectures using the same ID and corruption evaluation protocol as in the main paper.

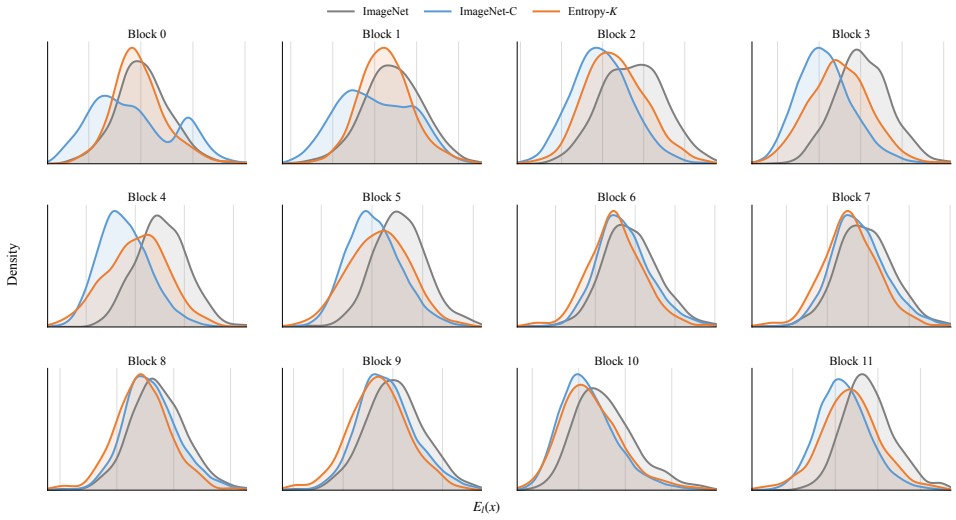
Additional PTQ baselines. The additional baselines represent several complementary directions in PTQ for Vision Transformers. PTQ4ViT introduces twin uniform quantization for post-Softmax and post-GELU activations, together with a Hessian-guided calibration metric. RepQ-ViT addresses inter-channel variation after LayerNorm and non-uniform post-Softmax activations through scale

reparameterization. NoisyQuant modifies activation distributions by adding a fixed noisy bias before quantization to mitigate heavy-tailed activations. AdaLog adopts adaptive logarithmic quantization for post-Softmax and post-GELU activations and searches for a suitable logarithmic base during calibration. I&S-ViT considers both the representation range of post-Softmax activations and the optimization stability of post-LayerNorm activations using a shift-uniform-log2 quantizer and a smooth optimization strategy. Together, these methods cover complementary approaches based on quantizer design, activation redistribution, reparameterization, and calibration optimization.

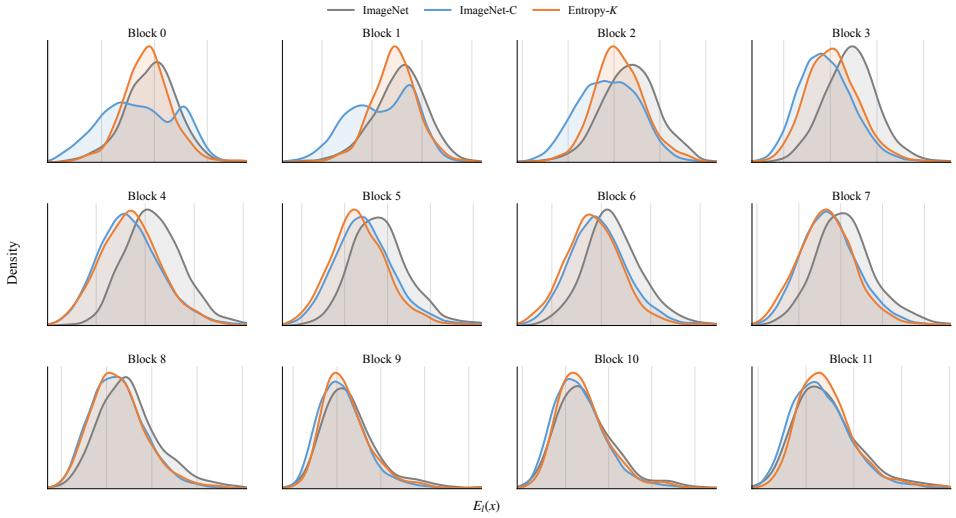
Extended quantitative comparison. The full results confirm the observations in the main paper when the comparison is extended to additional PTQ methods and the W2A4 setting. Under W3A3, Entropy-K achieves an Avg. of 57.14 across seven architectures, compared with 55.85 for the LS-ViT baseline. Under W2A4, it improves the Avg. from 52.09 to 53.11, while under the more aggressive W2A3 setting, the Avg. increases from 40.69 to 42.35. These results show that the benefit of transformation selection remains consistent across a broader range of quantization configurations. The per-architecture results further show that no single predefined transformation performs uniformly well across all models and precisions. In contrast, Entropy-K achieves the highest Avg. in all three quantization settings using the same entropy-selected transformation set without configuration-specific re-selection. Overall, the extended results support the consistency of the proposed selection strategy across architectures and bit-widths.

L Additional Visualization of Activation-Energy Distributions

Figure 10 visualizes activation-energy distributions across all Transformer blocks for ViT-S, DeiT-S, and Swin-S. Across architectures, ImageNet-C inputs generally exhibit lower activation energy than clean ImageNet inputs, with clearer shifts in the middle and later blocks. Entropy-K calibration samples show similar shifts toward the ImageNet-C distributions, particularly in these blocks. Although the degree of overlap varies across architectures and depth, the overall pattern is consistent with the main-text analysis.



(a) ViT-S



(b) DeiT-S



(c) Swin-S

Figure 10: **Activation-energy distributions across all Transformer blocks.** Distributions of clean ImageNet, ImageNet-C, and Entropy-K calibration inputs for the evaluated architectures.