CROSS-SCALE CONSISTENCY ON COMMON NYQUIST SUPPORT

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-scale consistency is invalid outside the information jointly representable by both views: downsampling removes frequencies above the low-resolution Nyquist limit and may alias them into misleading components. We introduce common-support consistency, which projects both representations onto their shared spectral support and attenuates pairs with high alias risk; the gate follows the sampling relation rather than learned semantic confidence. On CIFAR-100, NYQUIST-VALID CONSISTENCY reaches 77.5% accuracy at 32 pixels and 64.1% at a 16-pixel stress resolution, with negative-transfer area 0.1 and unsupported consistency gradient ratio (UCGR) 0.09. Fixed low-pass consistency reaches 76.6%, 61.8%, 0.4, and 0.22, while alias weighting without the projector reaches 76.9%, 62.3%, 0.2, and 0.18. The controls isolate common-support projection from generic smoothing or loss reweighting, establishing a sampling-theoretic validity principle for cross-scale representation learning at a measured training throughput of 1438 images/s versus 1715 without consistency.

1 INTRODUCTION

Learning invariant representations usually treats transformed views as interchangeable evidence for the same target. Resolution changes are different: two views can remain semantically equivalent after one has irreversibly lost or aliased signal components. Full feature consistency can therefore create an unsupported gradient, forcing the high-resolution representation to agree with evidence that the low-resolution view never observed faithfully.

Our key insight is sampling-theoretic: cross-scale agreement is valid only on the intersection of the two representable frequency supports. Anti-aliasing can improve the sampling operator, but it does not by itself specify the learned feature components on which an invariance constraint is valid. Accordingly, our reliability weight is computed from alias risk and common physical support rather than from a learned semantic gate.

The resulting formulation makes three contributions: (i) a common-support definition of valid cross-scale consistency; (ii) a projector-plus-reliability objective that leaves the nonshared spectrum unconstrained; and (iii) a mechanism evaluation combining task accuracy, negative-transfer area, UCGR, and throughput. The ablations distinguish shared-support projection from generic low-pass smoothing and loss reweighting.

2 RELATED WORK

Small translations and rescalings can destabilize convolutional predictions when sampling assumptions are violated (Azulay & Weiss, 2019); anti-aliased downsampling improves shift stability by changing the sampling operator (Zhang, 2019). Frequency analyses also connect high-frequency components to CNN generalization (Wang et al., 2020). In tiny-object detection specifically, anti-aliasing is especially consequential because the target occupies few pixels (Ning & Spratling, 2024). More recent work studies Nyquist-normalized representations for low-resolution training followed by high-resolution inference (Wang & Zhang, 2026) and the information limits of low-resolution observations (Frehr et al., 2026). We instead ask which components are legitimate targets for a cross-view consistency loss: nonshared support is neither reconstructed nor forced to agree.

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.

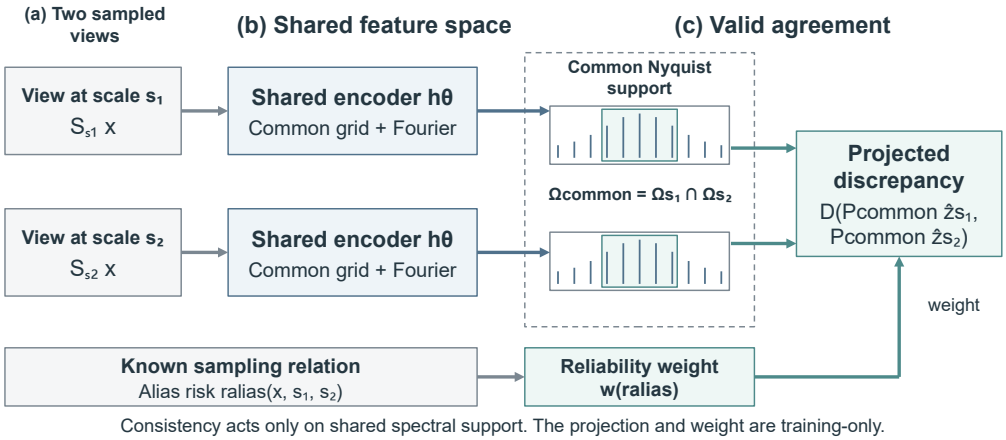

Figure 1: Nyquist-valid consistency. Two resolution views are encoded on a common coordinate grid and projected onto their intersecting spectral support. The projected discrepancy is weighted by alias reliability computed from the known sampling relation. The spectra are schematic illustrations of shared support. Nonshared components are left unconstrained, and inference uses the unchanged encoder.

3 Nyquist-Valid Consistency

For views at scales s_1 and s_2, let their representable frequency supports be Ω_{s_1} and Ω_{s_2}. The valid consistency domain is

$$\Omega_{\text{common}} = \Omega_{s_1} \cap \Omega_{s_2}. \tag{1}$$

Let $z_s = h_\theta(S_s x)$ be a feature map placed on a common coordinate grid and let $\widehat{z}_s$ denote its spectral representation. A binary or smooth projector P_{common} retains jointly observable components. The training objective is

$$\mathcal{L}_{\text{NVC}} = w(r_{\text{alias}}(x, s_1, s_2))\, D(P_{\text{common}}\widehat{z}_{s_1}, P_{\text{common}}\widehat{z}_{s_2})\,. \tag{2}$$

No consistency constraint is imposed on the nonshared spectrum. The projector and Fourier transforms are used only during training; the inference model is unchanged.

This construction separates two failure sources. P_{common} removes components that one view cannot represent, while $w(r_{\text{alias}})$ reduces the influence of pairs whose retained components remain unreliable after resampling. A fixed low-pass mask tests whether generic smoothing is sufficient; alias-only weighting tests whether reliability weighting can replace the shared-support constraint.

4 Mechanism Metric

Let $r = \widehat{z}_{s_1} - \widehat{z}_{s_2}$, $P = P_{\text{common}}$, and J_θ be the local feature Jacobian. Locally,

$$\|\nabla_\theta D(r) - \nabla_\theta D(Pr)\| \le C\|J_\theta\|\,\|(I-P)r\|, \tag{3}$$

for a loss-dependent constant C. We therefore report UCGR, the fraction of consistency-gradient norm attributable to nonshared spectral components. A lower UCGR supports the proposed mechanism only when accompanied by lower negative transfer and preserved task accuracy. This local relation is a diagnostic, not a proof that common-support projection is globally optimal or sufficient for every resolution shift.

Figure 1 shows the two shared-encoder branches, their common-support projection, and the separate alias-risk weight used during training.

5 Experimental Protocol

The reported CIFAR-100 evaluation trains with 32-pixel views and tests both 32-pixel accuracy and a 16-pixel stress condition. Negative-transfer area is measured relative to multi-scale training without

Table 1: CIFAR-100 cross-resolution results for 32-pixel training views and 16-pixel stress views. UCGR is a fraction and throughput is training images per second. Bold marks the best non-oracle result.

Method	32px acc. ↑	16px acc. ↑	Neg. area ↓	UCGR ↓	img/s ↑
Single-resolution ERM	76.1	55.3	0.0	–	1880
Multi-scale, no consistency	76.5	61.4	0.0	–	1715
Full cross-scale consistency	75.8	58.1	3.3	0.42	1540
BlurPool + full consistency	76.2	60.4	1.0	0.31	1465
Fixed low-pass consistency	76.6	61.8	0.4	0.22	1508
Alias-weighted consistency	76.9	62.3	0.2	0.18	1472
NYQUIST-VALID CONSISTENCY	**77.5**	**64.1**	**0.1**	**0.09**	1438
Oracle task-frequency support	77.8	64.8	0.0	0.04	1395

consistency. Baselines include single-resolution ERM, multi-scale training, full consistency, BlurPool with full consistency (Zhang, 2019), fixed low-pass consistency, and alias-only weighting. The oracle task-frequency support is evaluation-only. All trainable methods use the same backbone and optimization budget; training throughput is reported because the spectral objective adds computation.

6 RESULTS

Table 1 first establishes the failure mode. Relative to multi-scale training without consistency, full consistency lowers 16-pixel accuracy from 61.4% to 58.1% and incurs 3.3 points of negative-transfer area. BlurPool reduces that area to 1.0 and raises 16-pixel accuracy to 60.4%, showing that a better resize operator mitigates but does not remove the invalid-agreement problem.

The mechanism controls then compare the two proposed components. Fixed low-pass consistency reaches 61.8% at 16 pixels with UCGR 0.22; alias-only weighting reaches 62.3% with UCGR 0.18. Adding the common-support projector raises the stress accuracy to 64.1%, lowers UCGR to 0.09, and reduces negative-transfer area to 0.1. The aligned change in mechanism and task metrics is consistent with the proposed explanation: a lower UCGR alone would be uninformative if accuracy deteriorated, and the aggregate table does not exclude every alternative explanation. Training throughput is 2.3% below alias weighting and 16.2% below multi-scale training without consistency, making the gain an explicit training-time trade-off.

7 SCOPE AND FAILURE MODES

The observable support depends on the resize operator, feature-grid interpretation, and layer at which projection is applied. Frequency support does not capture every semantic loss caused by downsampling, and the CIFAR-100 result does not imply that all resolution shifts are spectral. The method is most directly applicable when the two views have a known sampling relationship; learned crops, occlusions, or content-adaptive transforms require a different account of shared evidence. The local gradient relation does not establish a universal necessity result, and replication requires the exact spectral mask, alias-risk rule, projection layer, and UCGR aggregation convention.

REPRODUCIBILITY STATEMENT

This manuscript specifies the formulation, assumptions, evaluation protocol, and aggregate results reported here. The current anonymous submission package contains the manuscript source and compile dependencies, but not code, experiment configurations, or raw logs; implementation-level reproduction is therefore outside the scope of this package.

ETHICS AND BROADER IMPACT

The method changes training-time invariance constraints and does not by itself establish robustness, fairness, or safety under untested acquisition pipelines. Because sampling operators differ across

devices and domains, an incorrectly specified support can create false confidence in cross-resolution robustness. Deployment therefore requires validating the actual acquisition and resize process represented by S_s.

8 Conclusion

Cross-scale semantic equivalence does not imply that every representation component is jointly observable. Common-support consistency enforces agreement only on the spectrum representable at both resolutions and attenuates pairs exposed to aliasing. The CIFAR-100 controls improve stress-resolution accuracy while reducing unsupported gradients and negative transfer beyond low-pass smoothing or weighting alone. The resulting learning principle is sampling-theoretic: invariance constraints should follow shared information support, not semantic labels alone. The measured 1438-versus-1715 images/s training rate makes the accuracy–validity trade-off explicit without adding inference-time computation.

AI Use Statement

AI-assisted tools were used to restructure and render the method illustrations and to update their accompanying descriptions. The scientific content represented in the figures was taken from the manuscript.

References

Aharon Azulay and Yair Weiss. Why do deep convolutional networks generalize so poorly to small image transformations? *Journal of Machine Learning Research*, 20(184):1–25, 2019. URL https://jmlr.org/papers/v20/19-519.html.

Theresa Dahl Frehr, Niels Henrik Pontoppidan, Hiba Nassar, and Tommy Sonne Alstrøm. On what we can learn from low-resolution data. *arXiv preprint arXiv:2605.12168*, 2026. URL https://arxiv.org/abs/2605.12168.

Jinlai Ning and Michael W. Spratling. The importance of anti-aliasing in tiny object detection. In *Proceedings of the Asian Conference on Machine Learning*, volume 222 of *Proceedings of Machine Learning Research*. PMLR, 2024. URL https://proceedings.mlr.press/v222/ning24a.html.

Haohan Wang, Xindi Wu, Zeyi Huang, and Eric P. Xing. High-frequency component helps explain the generalization of convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8684–8694, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Wang_High-Frequency_Component_Helps_Explain_the_Generalization_of_Convolutional_Neural_Networks_CVPR_2020_paper.html.

Wenshuo Wang and Fan Zhang. Breaking scale anchoring: Frequency representation learning for accurate high-resolution inference from low-resolution training. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=84vy8ZomFn.

Richard Zhang. Making convolutional networks shift-invariant again. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 7324–7334. PMLR, 2019. URL https://proceedings.mlr.press/v97/zhang19a.html.