Supplementary Material for "Rethinking Learning with Noisy Labels: A Critical Review of Structured Supervision, Representation Dynamics, and Robust-Learning Pipelines"

Wenxiao Fan[1] and Kan Li[1*]

[1*]School of Computer Science, Beijing Institute of Technology, No. 5 Zhongguancun South Street, Haidian District, Beijing, 100081, China.

*Corresponding author(s). E-mail(s): likan@bit.edu.cn;
Contributing authors: wenxiaofan@bit.edu.cn;

This supplementary document contains the supporting diagnostics and chronological evidence-status tables referenced by the main manuscript. Figure and table numbering continues from the main manuscript.

Appendix A Supporting Diagnostics for the Small-Loss Case Study

The main text retains the two figures and two tables needed to establish the central observability argument. This appendix provides four complementary checks that make the diagnostic chain explicit. The feature-space intervention asks whether corrupted examples remain high-loss when they are made easier to organize. The transition matrices show how the corruption mechanisms differ at the class level. The Gaussian mixture model (GMM) analysis separates aggregate splitting accuracy from clean-sample retention and noisy-sample detection. The final check replaces the loss-mixture decision with a confidence criterion and changes the downstream architecture. Together, these checks test different links in the argument rather than serving as additional method-comparison results.

All panels should be read within their displayed protocol. For synthetic corruption, the clean/noisy indicator is available from the injected corruption mask and is used only for retrospective diagnosis. It is not supplied to the learner as training supervision.

In the histogram panels, the horizontal axis is log loss and the two colors denote samples whose observed labels were retained or corrupted. In the curve panels, the horizontal axis is training epoch, while the legend suffix denotes the nominal corruption rate. The figures report the behavior of the displayed runs and are not used to claim statistical significance across independent implementations.

A.1 Controlled feature-space intervention

The first check isolates the mechanism behind the small-loss heuristic. Ordinarily, a corrupted target conflicts with the representation being learned and therefore tends to produce a larger loss. This association can be mistaken for a direct connection between loss and label correctness. The intervention reverses the diagnostic question: if corrupted examples are given feature-side structure that makes their observed targets easier to fit, do they remain distinguishable from clean examples by loss alone? CIFAR-10 with 40% instance-dependent noise provides the controlled setting, and the intervention is designed to increase the structural coherence of the corrupted group without changing which examples are counted as corrupted.

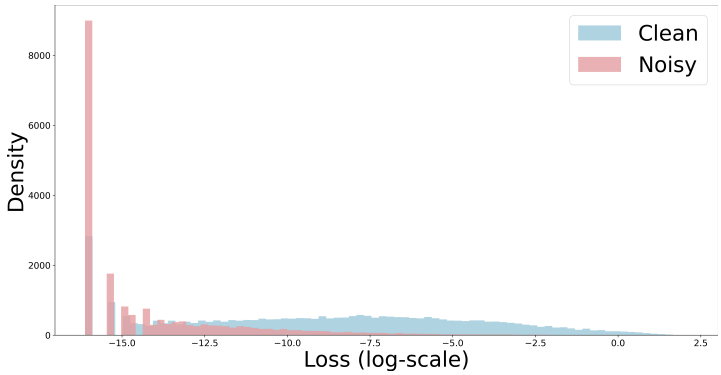
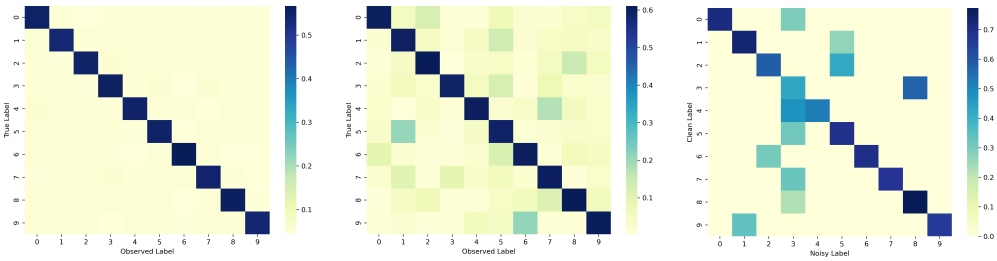


Fig. 7 Toy feature-space intervention on CIFAR-10 with 40% instance-dependent noise. The histograms show log-loss distributions for the known clean and corrupted groups after noisy samples are given clean-like feature structure. The resulting overlap shows that the small-loss signal depends on structural coherence, not only on label correctness.

Figure 7 shows the resulting loss distributions. A substantial fraction of corrupted examples moves into the same very-low-loss region occupied by clean examples. The noisy group has not become correctly labeled; it has become easier for the current model to fit under its observed targets. This distinction is the purpose of the intervention. It shows that a selector can assign high reliability to a corrupted example when representation geometry supports the corrupted target.

The intervention is intentionally narrow. It does not model a complete human annotation process, and it does not show that every structured corruption will become low-loss. It establishes a conditional failure mode: label correctness is not identifiable from loss ranking when clean and corrupted examples have comparable structural



Symmetric noise Instance-dependent noise SILN

Fig. 8 Empirical transition-matrix comparison on CIFAR-10. Rows denote clean classes and columns denote observed labels; diagonal cells represent retained labels and off-diagonal cells represent corruptions. Compared with broadly distributed symmetric and instance-dependent transitions, SILN concentrates substantial mass on selected class pairs.

learnability. This motivates inspecting the corruption pattern before interpreting a favorable small-loss separation as a general property of an algorithm.

A.2 Class-level structure of the corruption process

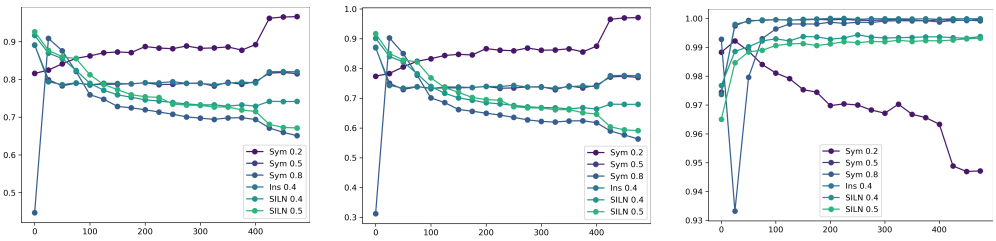
The transition matrices make the three synthetic protocols comparable at the class level. Each row corresponds to a clean class and each column to the observed class after corruption. Diagonal mass represents label retention, whereas off-diagonal mass records where corrupted labels are sent. A matrix with diffuse off-diagonal mass creates a different learning problem from one in which a class is redirected repeatedly toward a small number of alternatives, even when the two protocols have a similar global corruption rate.

Figure 8 visualizes this distinction. Symmetric noise preserves a dominant diagonal and distributes the remaining mass broadly. Instance-dependent noise produces heterogeneous off-diagonal entries because the corruption depends on individual examples. SILN is more concentrated: selected class pairs receive visibly stronger transition mass, so the corrupted targets can form repeatable class-level patterns rather than isolated random contradictions.

The matrix view explains why noise rate alone is insufficient for this case study. Two datasets can contain the same proportion of changed labels while exposing very different reliability signals to a learner. Diffuse corruptions are more likely to conflict with shared class structure, whereas concentrated transitions can be absorbed as an alternative regularity. The figure is therefore used to document the stress-test structure, not to claim that SILN reproduces every real annotation mechanism.

A.3 GMM splitting and hard-clean retention

The third check evaluates the decision rule used by many small-loss pipelines. At each displayed training point, a two-component GMM summarizes the loss distribution and converts the mixture assignment into a clean/noisy decision. Because the corruption mask is known for this controlled analysis, three quantities can be reported: overall agreement with the mask, the fraction of clean samples retained as clean, and the



Overall splitting accuracy Clean-sample accuracy Noisy-sample accuracy

Fig. 9 GMM clean/noisy splitting accuracy across training under RoLR on CIFAR-100. The panels report overall agreement with the injected corruption mask, clean-sample retention, and noisy-sample detection. Comparing the panels reveals whether an apparently strong aggregate split is achieved by sacrificing hard clean samples.

fraction of corrupted samples identified as noisy. The second and third quantities are important because the aggregate score can conceal an asymmetric error pattern.

Figure 9 reveals such an asymmetry. Noisy-sample detection remains close to the top of its plotting range for most displayed settings, but clean-sample retention varies substantially and decreases in several high-noise or structured conditions. Consequently, the overall curve can look comparatively stable even while a growing portion of clean data is assigned to the noisy component. This is precisely the failure that a single aggregate splitting score can hide.

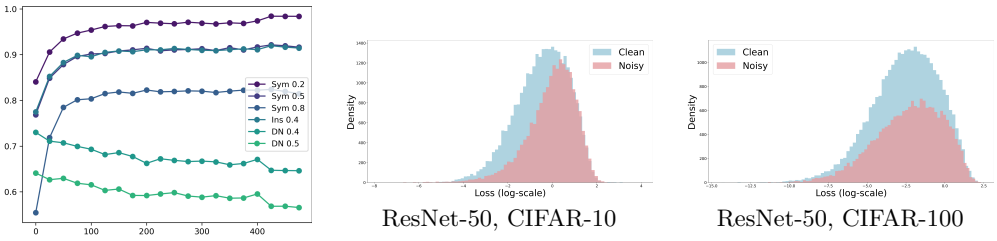
The clean-sample panel should therefore be interpreted as a hard-clean retention diagnostic. A clean example can remain high-loss because it lies near a class boundary, belongs to a difficult subclass, or is poorly represented at the current epoch. Rejecting such examples may improve the apparent purity of the retained set while reducing coverage of informative training data. The noisy-sample panel measures the complementary error. Reporting both prevents a method from appearing successful merely because it applies an increasingly conservative selection rule.

The curves also show why threshold retuning is not a complete remedy. A different operating point can trade clean retention against noisy detection, but it cannot restore separability when the underlying score distributions overlap. The relevant question is therefore not only which threshold is used, but whether loss contains enough information for the desired clean/noisy decision at that stage of training.

A.4 Alternative score and architecture-transfer checks

The final group addresses two narrower alternative explanations. First, poor separation might arise from the two-component GMM rather than from the underlying reliability signal. The confidence panel therefore applies a high-confidence criterion under the same RoLR-centered diagnostic setting. This changes the score used for selection while preserving the question being evaluated: whether clean and corrupted samples can be separated without systematically rejecting difficult clean examples.

Second, the observed overlap might be circular if the same model both constructs the structured noisy labels and evaluates them. The architecture-transfer panels reduce this concern by using ResNet-50 for downstream evaluation when the SILN labels



Confidence criterion ResNet-50, CIFAR-10 ResNet-50, CIFAR-100

Fig. 10 Alternative-score and architecture-transfer checks. The left panel applies a high-confidence criterion under RoLR on CIFAR-100. The middle and right panels show clean/noisy log-loss distributions when SILN labels generated with one model are evaluated with ResNet-50 on CIFAR-10 and CIFAR-100.

were generated with another model. The two histograms repeat the clean/noisy log-loss comparison on CIFAR-10 and CIFAR-100, allowing the transfer check to be read across datasets as well as across model instances.

Figure 10 shows that changing the decision score does not automatically remove the clean/noisy trade-off, and changing the downstream architecture does not eliminate loss overlap in the displayed settings. These observations make the diagnostic less dependent on one particular GMM fit or one generator–evaluator pairing. They do not establish architecture-independent behavior in general: only one transfer architecture and the two displayed datasets are tested here.

A.5 Interpretation boundary

Taken together, the appendix supports three limited conclusions. First, small loss is a model- and representation-dependent reliability signal rather than a direct observation of label correctness. Second, aggregate selection accuracy can hide the removal of hard clean examples, so clean retention and noisy detection should be reported separately. Third, the observed difficulty is not confined to one mixture model or one identical-model construction, because it remains visible under the displayed confidence and architecture-transfer checks.

The appendix does not establish that SILN is a universal model of realistic label noise, that every small-loss method fails under structured corruption, or that the displayed runs rank competing algorithms. The evidence is diagnostic and protocol-specific. Its practical implication is a reporting requirement: an LNL evaluation should pair final prediction accuracy with the corruption structure, score distributions, group-specific selection errors, and the information used to construct or validate the noisy labels. These additions make clear whether a method has learned a robust predictor or has merely benefited from an unusually favorable clean/noisy separation.

Appendix B Evidence-Status Tables for Recent Literature

The chronological tables below preserve publication status and provenance without making publication year the organizing axis of the main review. The 2026 entries are

Table 17 Representative studies related to noisy labels or noisy supervision, 2023–2024.

Paper	Venue/year	Main setting	Role in this review
DISC (Li et al. 2023)	CVPR 2023	Closed-set LNL	Instance-specific memorization and dynamic selection/correction
Fine-grained LNL (Wei et al. 2023b)	CVPR 2023	Fine-grained classification	Semantic ambiguity and representation-preserving contrastive learning
Noisy correspondence (Han et al. 2023)	CVPR 2023	Cross-modal retrieval	Noisy supervision as mismatched correspondence
RankMatch (Zhang et al. 2023)	ICCV 2023	Sample selection	Confidence and rank consistency beyond small-loss selection
Intrinsically long-tailed LNL (Lu et al. 2023)	ICCV 2023	Long-tailed LNL	Coupling between imbalance, difficulty, and label reliability
Method-choice analysis (Yao et al. 2023)	ICML 2023	Method assumptions	Causal assumptions behind SSL-style and noise-modeling approaches
LogitClip (Wei et al. 2023a)	ICML 2023	Robust objective	Optimization-level control of memorization
CSOT (Chang et al. 2023)	NeurIPS 2023	Structured selection/correction	Global and local structure-aware denoising
FedNoRo (Wu et al. 2023)	IJCAI 2023	Federated LNL	Client heterogeneity, class imbalance, and noisy labels
Noisy pre-training (Chen et al. 2024, 2025)	ICLR 2024 / TPAMI 2025	Noisy model learning	Label noise embedded in pre-trained representations and foundation-model adaptation
Structural labels (Kim et al. 2024)	CVPR 2024	Structure-aware LNL	Reverse kNN structural targets for representation geometry
L2B (Zhou et al. 2024)	CVPR 2024	Bootstrapping	Joint instance and label weighting for robust self-bootstrapping
VLM noisy-label detector (Wei et al. 2024)	NeurIPS 2024	VLM fine-tuning	External semantic priors for noisy-label detection

used as pressure tests rather than as the sole support for core claims. Peer-reviewed work is separated from preprint and emerging evidence so that the appendix can be updated as publication records change.

Table 18 Representative studies related to noisy labels or noisy supervision, 2025.

Paper	Venue/year	Main setting	Role in this review
Open-world noisy data (Pan et al. 2025)	AAAI 2025	Open-world LNL	Class-independent margins for known/unknown noise
Multi-prototype open-set LNL (Zhang et al. 2025)	Pattern Recognition 2025	Open-set LNL	Feature-prototype modeling for clean/ID-noisy/OOD-noisy separation
ANNE (Cordeiro and Carneiro 2025) and surrogate selection (Liang et al. 2025)	Pattern Recognition / IJCV 2025	Sample selection	Hybrid loss-feature selection and general reliability surrogates
VRI (Sun et al. 2025) and Detect-and-Correct (Grinberg et al. 2025)	IJCV / arXiv 2025	Target refinement	Probabilistic and selective correction under uncertain targets
SiDyP (Ye et al. 2025)	KDD 2025	LLM-generated labels	LLM annotation noise and iterative target refinement
BeGIN (Kim et al. 2025)	KDD 2025	Graph benchmarks	Instance-dependent graph label noise with algorithmic and LLM-based corruptions
Early Cutting (Yuan et al. 2025)	NeurIPS 2025	Sample selection	Filtering mislabeled easy examples after early confidence
Joint asymmetric loss (Wang et al. 2025) and RML++ (Li et al. 2025)	ICCV / IJCV 2025	Robust loss	Robust objective design beyond symmetric losses
ELDET (Choi et al. 2025)	NeurIPS 2025	Object detection	Early-learning distillation under localization and categorization noise

Table 19 Peer-reviewed 2026 evidence in noisy-label learning.

Paper	Evidence status	Main setting	Role in this review
Label-noise SGD dynamics (Zhang et al. 2026)	Published, AAAI 2026	Theory and optimization	Explains how label noise can drive rich-regime dynamics and connects noisy-label training to SAM-like optimization
Jump-teaching (Ji et al. 2026)	Published, AAAI 2026	Efficient sample selection	Uses temporal disagreement to reduce selection bias without dual networks or extra forward passes
DKAF (Ning and Chen 2026)	Published, AAAI 2026	VLM noisy-label learning	Shows that VLM priors may create endogenous confirmation bias during noisy-label detection and correction
FedRNC (Huang et al. 2026)	Published, AAAI 2026	Federated class-incremental LNL	Introduces spatio-temporal label misalignment as a federated noisy-supervision failure mode
Prototype-guided graph supervision (Li et al. 2026a)	Published, AAAI 2026	Graph learning	Replaces direct noisy node labels with class-level prototype supervision under sparse labels

Table 20 Foundation-model, conformal, and emerging 2026 evidence.

Paper	Evidence status	Main setting	Role in this review
LANE (Sosea and Caragea 2026)	Published, ICLR 2026	Fine-grained text classification	Uses label relationships and label-aware margins to detect semantic label noise in text
RRM (Chen et al. 2026)	Published, ICLR 2026	Loss reweighting	Provides an architecture-independent wrapper for loss reweighting under noisy labels
Noise-aware generalization (Wang et al. 2026)	Published, ICLR 2026	Domain generalization plus noise	Shows that LNL methods may confuse domain shift with label noise, motivating joint evaluation
RE-PO (Cao et al. 2026) and Semi-DPO (Liu et al. 2026)	Published, ICLR 2026	Preference/noisy alignment labels	Extends noisy-supervision reasoning to LLM and diffusion preference optimization
CMRM (Shi et al. 2026)	Published, AISTATS 2026	Conformal robust learning	Adds a quantile-calibrated margin envelope without privileged clean data or transition matrices
NCSAM (Xu and Pang 2026) and OccNL (Li et al. 2026b)	NCSAM: arXiv preprint; OccNL: IROS 2026 acceptance reported by the authors	Optimization and 3D perception	Indicate emerging work on flatness-aware noisy optimization and task-specific 3D occupancy noise

References

Cao X, Xu Z, Guang M, et al (2026) RE-PO: Robust enhanced policy optimization as a general framework for LLM alignment. In: International Conference on Learning Representations, URL https://openreview.net/forum?id=jDKpOvTCM8

Chang W, Shi Y, Wang J (2023) CSOT: Curriculum and structure-aware optimal transport for learning with noisy labels. In: Advances in Neural Information Processing Systems, pp 8528–8541, https://doi.org/10.52202/075280-0373, URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/1b0da24d136f46bfaee78e8da907127e-Abstract-Conference.html

Chen H, Wang J, Shah AP, et al (2024) Understanding and mitigating the label noise in pre-training on downstream tasks. In: International Conference on Learning Representations, URL https://openreview.net/forum?id=TjhUtloBZU

Chen H, Wang Z, Tao R, et al (2025) Impact of noisy supervision in foundation model learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 47(7):5690–5707. https://doi.org/10.1109/TPAMI.2025.3552309, URL https://ieeexplore.ieee.org/document/10934976

Chen L, Chern B, Eckstrand E, et al (2026) Enhancing learning with noisy labels via rockafellian relaxation. In: International Conference on Learning Representations, URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/7e810b2c75d69be186cadd2fe3febeab-Abstract-Conference.html

Choi D, Lee S, Yun E, et al (2025) ELDET: Early-learning distillation with noisy labels for object detection. In: Advances in Neural Information Processing Systems, pp 69345–69372, https://doi.org/10.52202/085713-2331, URL https://papers.neurips.cc/paper_files/paper/2025/hash/6460e378f24da3a79f20ac2640732a00-Abstract-Conference.html

Cordeiro FR, Carneiro G (2025) ANNE: Adaptive nearest neighbours and eigenvector-based sample selection for robust learning with noisy labels. Pattern Recognition 159:111132. https://doi.org/10.1016/j.patcog.2024.111132, URL https://www.sciencedirect.com/science/article/pii/S0031320324008835

Grinberg Y, Harel N, Goldberger J, et al (2025) Detect and correct: A selective noise correction method for learning with noisy labels. Computer Science Research Notes 3501(2025):173–181. https://doi.org/10.24132/CSRN.2025-19, URL https://doi.org/10.24132/CSRN.2025-19, 33rd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2025)

Han H, Miao K, Zheng Q, et al (2023) Noisy correspondence learning with meta similarity correction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 7517–7526, https://doi.org/10.1109/CVPR52729.2023.00726, URL https://openaccess.thecvf.

com/content/CVPR2023/html/Han_Noisy_Correspondence_Learning_With_Meta_Similarity_Correction_CVPR_2023_paper.html

Huang X, Sun Z, Shi J, et al (2026) FedRNC: Addressing spatio-temporal label misalignment in federated noisy class-incremental learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 22048–22056, https://doi.org/10.1609/aaai.v40i26.39359, URL https://ojs.aaai.org/index.php/AAAI/article/view/39359

Ji K, Cheng F, Wang Z, et al (2026) Jump-teaching: Combating sample selection bias via temporal disagreement. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 22191–22199, https://doi.org/10.1609/aaai.v40i26.39375, URL https://ojs.aaai.org/index.php/AAAI/article/view/39375

Kim Nr, Lee JS, Lee JH (2024) Learning with structural labels for learning with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 27610–27620, https://doi.org/10.1109/CVPR52733.2024.02607, URL https://openaccess.thecvf.com/content/CVPR2024/html/Kim_Learning_with_Structural_Labels_for_Learning_with_Noisy_Labels_CVPR_2024_paper.html

Kim S, Kang SK, Kim D, et al (2025) Delving into instance-dependent label noise in graph data: A comprehensive study and benchmark. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, pp 5539–5550, https://doi.org/10.1145/3711896.3737376, URL https://doi.org/10.1145/3711896.3737376

Li F, Li K, Wang Q, et al (2025) RML++: Regroup median loss for combating label noise. International Journal of Computer Vision 133(9):6400–6421. https://doi.org/10.1007/s11263-025-02494-4, URL https://doi.org/10.1007/s11263-025-02494-4

Li Q, Li X, Li D, et al (2026a) Prototype-guided supervision for graph learning with noisy and sparse labels. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 23105–23113, https://doi.org/10.1609/aaai.v40i27.39477, URL https://ojs.aaai.org/index.php/AAAI/article/view/39477

Li W, Peng K, Wen D, et al (2026b) Can we trust unreliable voxels? exploring 3d semantic occupancy prediction under label noise. https://doi.org/10.48550/arXiv.2603.06279, URL https://arxiv.org/abs/2603.06279, accepted to IROS 2026, arXiv:2603.06279

Li Y, Han H, Shan S, et al (2023) DISC: Learning from noisy labels via dynamic instance-specific selection and correction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 24070–24079, https://doi.org/10.1109/CVPR52729.2023.02305, URL https://openaccess.thecvf.com/content/CVPR2023/html/Li_DISC_Learning_From_Noisy_Labels_via_Dynamic_Instance-Specific_Selection_and_CVPR_2023_paper.html

Liang C, Zhu L, Shi H, et al (2025) Combating label noise with a general surrogate model for sample selection. International Journal of Computer Vision 133(6):3166–3179. https://doi.org/10.1007/s11263-024-02324-z, URL https://doi.org/10.1007/s11263-024-02324-z

Liu X, Li M, Lyu Z, et al (2026) Learning from noisy preferences: A semi-supervised learning approach to direct preference optimization. In: International Conference on Learning Representations, pp 13690–13709, URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/17061a94c3c7fda5fa24bbdd1832fa99-Abstract-Conference.html

Lu Y, Zhang Y, Han B, et al (2023) Label-noise learning with intrinsically long-tailed data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 1369–1378, https://doi.org/10.1109/ICCV51070.2023.00132, URL https://openaccess.thecvf.com/content/ICCV2023/html/Lu_Label-Noise_Learning_with_Intrinsically_Long-Tailed_Data_ICCV_2023_paper.html

Ning F, Chen X (2026) Mitigating endogenous confirmation bias in noisy label learning for vision-language models. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 24576–24584, https://doi.org/10.1609/aaai.v40i29.39641, URL https://ojs.aaai.org/index.php/AAAI/article/view/39641

Pan L, Gao C, Zhou J, et al (2025) Learning with open-world noisy data via class-independent margin in dual representation space. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 6290–6298, https://doi.org/10.1609/aaai.v39i6.32673

Shi Y, Li P, Zhang Z, et al (2026) Conformal margin risk minimization: An envelope framework for robust learning under label noise. In: Proceedings of The 29th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol 300. PMLR, pp 4276–4284, URL https://proceedings.mlr.press/v300/shi26b.html, arXiv:2604.06468

Sosea T, Caragea C (2026) LANE: Label-aware noise elimination for fine-grained text classification. In: International Conference on Learning Representations, URL https://openreview.net/forum?id=PuayLKdrFz

Sun H, Wei Q, Feng L, et al (2025) Variational rectification inference for learning with noisy labels. International Journal of Computer Vision 133(2):652–671. https://doi.org/10.1007/s11263-024-02205-5, URL https://doi.org/10.1007/s11263-024-02205-5

Wang J, Liu X, Zhou X, et al (2025) Joint asymmetric loss for learning with noisy labels. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 1947–1956, https://doi.org/10.1109/ICCV51701.2025.00189, URL https://openaccess.thecvf.com/content/ICCV2025/html/Wang_Joint_Asymmetric_Loss_for_Learning_with_Noisy_Labels_ICCV_2025_paper.html

Wang S, Liu A, Plummer BA (2026) Noise-aware generalization: Robustness to in-domain noise and out-of-domain generalization. In: International Conference on Learning Representations, URL https://openreview.net/forum?id=wb83wO41QT

Wei H, Zhuang H, Xie R, et al (2023a) Mitigating memorization of noisy labels by clipping the model prediction. In: Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol 202. PMLR, pp 36868–36886, URL https://proceedings.mlr.press/v202/wei23e.html

Wei Q, Feng L, Sun H, et al (2023b) Fine-grained classification with noisy labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 11651–11660, https://doi.org/10.1109/CVPR52729.2023.01121, URL https://openaccess.thecvf.com/content/CVPR2023/html/Wei_Fine-Grained_Classification_With_Noisy_Labels_CVPR_2023_paper.html

Wei T, Li HT, Li CS, et al (2024) Vision-language models are strong noisy label detectors. In: Advances in Neural Information Processing Systems, pp 58154–58173, https://doi.org/10.52202/079017-1854, URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/6af08ba9468f0daca4b8dd388cb95824-Abstract-Conference.html

Wu N, Yu L, Jiang X, et al (2023) FedNoRo: Towards noise-robust federated learning by addressing class imbalance and label noise heterogeneity. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, pp 4424–4432, https://doi.org/10.24963/ijcai.2023/492, URL https://www.ijcai.org/proceedings/2023/492

Xu J, Pang J (2026) NCSAM noise-compensated sharpness-aware minimization for noisy label learning. https://doi.org/10.48550/arXiv.2601.19947, URL https://arxiv.org/abs/2601.19947, arXiv:2601.19947

Yao Y, Gong M, Du Y, et al (2023) Which is better for learning with noisy labels: The semi-supervised method or modeling label noise? In: Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol 202. PMLR, pp 39660–39673, URL https://proceedings.mlr.press/v202/yao23a.html

Ye L, Shah A, Zhang C, et al (2025) Calibrating pre-trained language classifiers on LLM-generated noisy labels via iterative refinement. In: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2. Association for Computing Machinery, pp 3598–3609, https://doi.org/10.1145/3711896.3736871

Yuan S, Feng L, Han B, et al (2025) Enhancing sample selection against label noise by cutting mislabeled easy examples. In: Advances in Neural Information Processing Systems, pp 49163–49191, https://doi.org/10.52202/085713-1466, URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/

3e826f682178d9830a3b704141b4989e-Abstract-Conference.html

Zhang T, Zhou Z, Wang M, et al (2026) On the learning dynamics of two-layer linear networks with label noise SGD. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 28400–28408, https://doi.org/10.1609/aaai.v40i33.40069, URL https://ojs.aaai.org/index.php/AAAI/article/view/40069

Zhang Y, Chen Y, Fang C, et al (2025) Learning from open-set noisy labels based on multi-prototype modeling. Pattern Recognition 157:110902. https://doi.org/10.1016/j.patcog.2024.110902, URL https://www.sciencedirect.com/science/article/pii/S0031320324006538

Zhang Z, Chen W, Fang C, et al (2023) RankMatch: Fostering confidence and consistency in learning with noisy labels. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 1644–1654, https://doi.org/10.1109/ICCV51070.2023.00158, URL https://openaccess.thecvf.com/content/ICCV2023/html/Zhang_RankMatch_Fostering_Confidence_and_Consistency_in_Learning_with_Noisy_Labels_ICCV_2023_paper.html

Zhou Y, Li X, Liu F, et al (2024) L2B: Learning to bootstrap robust models for combating label noise. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 23523–23533, https://doi.org/10.1109/CVPR52733.2024.02220, URL https://openaccess.thecvf.com/content/CVPR2024/html/Zhou_L2B_Learning_to_Bootstrap_Robust_Models_for_Combating_Label_Noise_CVPR_2024_paper.html