ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2026.0569

CAP-YOLO: TEXTURE RESIDUAL PRIOR ENCODING FOR LIGHTWEIGHT YOLO-BASED FABRIC DEFECT DETECTION

Maizhe Jiang[1] and Jihao Liu[2]
[1]*School of Electrical and Electronic Engineering, Wenzhou University, China*
[2]*School of Mechanical and Electrical Engineering, Jinggangshan University, China*

Abstract
Fabric defect detection remains difficult when defects appear as subtle structural disruptions rather than strong color changes. This paper presents CAP-YOLO, a lightweight prior-coded input framework for YOLO-based fabric defect detection. Instead of redesigning the detector architecture, the method converts image-derived priors into RGB-compatible inputs so that multiple variants can be compared under the same YOLO26n detector, parameter count, training schedule, and data split. Two image-only priors are studied: a Delta E color-deviation prior and an anomaly-derived texture residual prior. Five variants are evaluated on the Tianchi fabric defect dataset: an RGB baseline (M0), Delta E prior coding (M1), anomaly prior coding (M2), dual-prior coding (M3), and a lightweight prior-adapter preprocessing route (M4). On the held-out test split, M2 achieves the best overall result with 0.3289 mAP50 and 0.1542 mAP50-95, improving over the RGB baseline by 18.48% in mAP50, 27.26% in precision, 9.38% in recall, and 16.62% in F1 score, while keeping the same detector size of 2.3787M parameters. Per-class analysis shows that the gains are concentrated in texture-disruptive defect categories, whereas categories dominated by diffuse appearance or subtle color spread benefit less. These results support a practical conclusion: for lightweight fabric defect detection, a simple image-derived texture residual prior is more effective than a color-only prior or naive multi-prior mixing under the current controlled setting.

Keywords:
Fabric Defect Detection, YOLO, Texture Residual Prior, Prior-Coded RGB Input, Industrial Image Processing

1. INTRODUCTION

Automated fabric defect detection is an important requirement in textile manufacturing because missed defects directly affect product quality, inspection efficiency, and downstream cost. Although modern object detectors provide a useful balance between speed and accuracy, many fabric defects remain challenging because they are small, sparse, and visually subtle. In practical inspection images, defects often appear as local structural disturbances, broken weave regularity, or surface disruption rather than as strong color deviations. Under such conditions, an RGB-only detector may fail to emphasize the most relevant defect cues [1]-[3].

One practical way to improve detector input quality is to inject domain-relevant prior information before detection. In fabric inspection, two intuitive prior types are color-oriented cues and texture-oriented cues. A color difference prior can highlight localized chromatic deviation, while a texture residual prior can highlight disruption of repeated local structure. However, not every prior is equally useful, and simply combining multiple priors does not guarantee additive benefit. Therefore, before designing a stronger fusion architecture, it is necessary to determine which prior is actually helpful under a controlled experimental setting.

This paper studies that question through CAP-YOLO, a prior-coded input framework for lightweight YOLO-based fabric defect detection. The core idea is intentionally conservative: instead of modifying the detector backbone, image-derived prior maps are encoded into standard three-channel inputs so that multiple variants can be evaluated under the same detector architecture. This strategy keeps the detector unchanged and makes the comparison easier to interpret. All formal experiments use the same YOLO26n detector, data split, image size, and 50-epoch training schedule.

The study evaluates five variants. M0 is the RGB baseline. M1 injects a Delta E color prior. M2 injects an anomaly-derived texture residual prior. M3 mixes both priors with fixed coding. M4 uses a lightweight prior-adapter preprocessing route that generates pseudo-RGB inputs before standard YOLO training. The central result is that M2, the anomaly-prior-only variant, is the strongest method on the held-out test split. This result is notable not because it introduces a radically new detector, but because it shows that a simple and reproducible image-derived texture prior can produce a consistent gain without modifying the lightweight YOLO backbone.

The main contributions of this work are as follows: (1) A lightweight prior-coded input framework is established for fabric defect detection without modifying the YOLO26n detector architecture. (2) Two image-derived priors, Delta E color deviation and anomaly-based texture residual, are compared under a controlled M0-M4 ablation protocol. (3) Formal held-out test-set results show that the texture residual prior is the strongest tested variant, outperforming the RGB baseline and the other prior-coding routes. (4) Per-class and qualitative analyses identify where the texture prior helps most and where its current formulation remains limited.

2. RELATED WORK

2.1 FABRIC DEFECT DETECTION

Fabric defect detection has evolved from handcrafted image processing and texture analysis to deep-learning-based classification, segmentation, and detection methods. Traditional approaches based on filtering, frequency analysis, local texture operators, or adaptive thresholding can work in restricted environments, but they often degrade when defect appearance becomes diverse and background texture becomes complex. Deep neural networks improve robustness and representation power, but their performance still depends heavily on how defect cues appear in the detector input [1]-[3].

2.2 LIGHTWEIGHT OBJECT DETECTION FOR INDUSTRIAL INSPECTION

YOLO-style detectors are widely used in industrial inspection because they provide near-real-time performance with end-to-end localization and classification. In practical deployment scenarios, lightweight detectors are particularly attractive because inspection systems often run on constrained hardware or high-throughput production lines. This motivates keeping the detector compact while improving the input representation, rather than relying solely on larger backbones [4]-[6]. Classical two-stage and one-stage detectors also provide important background for interpreting lightweight detection tradeoffs [7]-[10].

2.3 PRIOR-GUIDED AND ANOMALY-INSPIRED VISUAL PROCESSING

Auxiliary priors have long been used to strengthen visual processing pipelines. In industrial inspection, anomaly-related cues are especially relevant because they describe deviation from expected structure. Recent anomaly detection research shows that residuals, reconstruction errors, and local irregularity maps can highlight subtle surface abnormalities. However, most of those methods are designed for anomaly segmentation or anomaly classification rather than object detection. In contrast, the present work uses an image-derived prior only as an input enhancement for a standard detector. The goal is not to train a separate anomaly model, but to test whether a lightweight, label-free prior can improve defect detection when encoded into RGB-compatible inputs [11]-[13]. Industrial anomaly-detection benchmarks and patch-level modeling methods further support the relevance of local deviation cues for visual inspection [14]-[16].

3. METHOD

3.1 OVERVIEW

CAP-YOLO is a prior-coded input strategy rather than a new detector backbone. Given an RGB fabric image, the pipeline computes one or more image-derived prior maps and encodes them into a standard three-channel image representation. The resulting image is then used to train and evaluate an unmodified YOLO26n detector [6]. This design keeps all variants comparable at the detector level and isolates the effect of prior coding. The overall pipeline is illustrated in Fig.1.

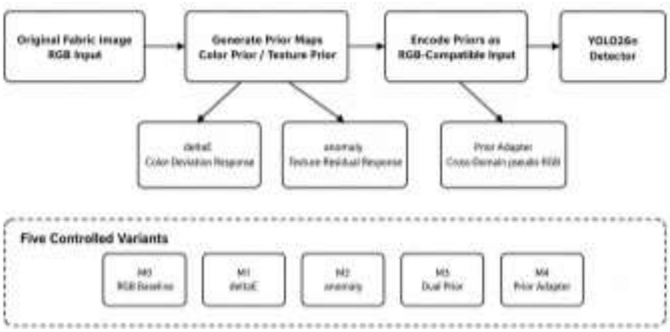

Fig.1. Overall prior-coded YOLO pipeline

The five evaluated variants are summarized in Table 1.

Table.1. Formal Ablation Variants

Variant	Input construction	Detector
M0	Original RGB image	YOLO26n
M1	RGB coding with Delta E prior	YOLO26n
M2	RGB coding with anomaly prior	YOLO26n
M3	RGB coding with Delta E and anomaly priors	YOLO26n
M4	Offline adapter-style pseudo-RGB preprocessing	YOLO26n

3.2 DELTA E COLOR PRIOR

The Delta E prior is designed to emphasize chromatic deviation from the dominant appearance of the current image. The RGB image is converted into CIE Lab color space, and a median color reference is computed from the image. For each pixel, the Euclidean distance from the pixel color to the image-level median is measured. The resulting map is clipped by a high percentile and normalized into an 8-bit single-channel image. This prior is entirely image-derived and does not use labels, boxes, or masks. The use of Delta E follows established color-difference research in CIE-based color spaces [17].

3.3 ANOMALY-DERIVED TEXTURE RESIDUAL PRIOR

The anomaly prior is designed to highlight local structural irregularity. The image is first converted to grayscale. A smoothed background is estimated with median filtering, and an absolute residual is computed between the original grayscale image and the smoothed background. A local texture-variation term is then combined with the residual term to produce the final anomaly map. After normalization, the map emphasizes localized discontinuities, structural disruption, and departures from repeated weave texture. Like the Delta E prior, this map is fully image-derived and avoids label leakage.

3.4 RGB-COMPATIBLE PRIOR CODING

For M1-M3, the prior maps are encoded into a standard RGB image using fixed alpha blending. The purpose is to preserve compatibility with the standard YOLO training pipeline while exposing the detector to additional prior cues. M1 encodes the Delta E map into the red channel. M2 encodes the anomaly prior into the green channel. M3 encodes both priors simultaneously into the red and green channels. In all three cases, the output remains a three-channel image.

M4 follows a different route. It generates a pseudo-RGB image through a lightweight prior-adapter preprocessing rule that injects small prior components into the original channels before standard YOLO training. Importantly, the M4 result reported in this paper corresponds to an offline preprocessing route rather than a separately validated end-to-end learnable fusion module. Therefore, M4 should be interpreted as an adapter-style preprocessing baseline, not as proof of trainable fusion.

4. EXPERIMENTAL SETUP

4.1 DATASET

Experiments were conducted on the Tianchi fabric defect dataset converted into YOLO format. The full dataset contains 9,576 images and corresponding annotations [18]. The split used in this study contains 6,702 training images, 1,913 validation images, and 961 held-out test images covering 21 defect categories.

4.2 CONTROLLED PROTOCOL

All five variants were trained under the same protocol: detector YOLO26n [6], parameters 2.3787M, image size 640, batch size 64, epochs 50, seed 42, and the same train, validation, and test split for all variants. This shared configuration allows the comparison to focus on the effect of input prior coding rather than detector scaling or training inconsistency. The dataset split and controlled experimental routes are summarized in Fig.2. The detector implementation follows the Ultralytics YOLO training interface [6].

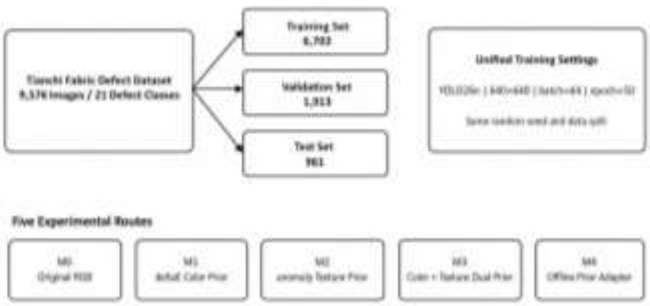

Fig.2. Dataset split and controlled experimental routes.

4.3 EVALUATION METRICS

The study reports precision, recall, mAP50, mAP50-95, F1 score, approximate FPS, and parameter count. Validation metrics are used as training-stage ablation evidence, while the final comparative claims in this paper are based on a separate formal evaluation on the held-out test split using the best checkpoint from each variant. Because the detector architecture is identical across all five variants, parameter count remains constant at 2.3787M. The reported FPS values are relative efficiency indicators obtained under the same evaluation environment and should not be interpreted as deployment-certified throughput numbers.

5. RESULTS AND DISCUSSION

5.1 MAIN HELD-OUT TEST RESULTS

Table 2 presents the formal held-out test-set results for all five variants. M2 is the strongest overall, achieving the best precision, recall, mAP50, mAP50-95, and F1 score among all configurations.

Relative to the RGB-only baseline M0, M2 improves mAP50 by 18.48%, precision by 27.26%, recall by 9.38%, and F1 by 16.62%. This result provides the clearest support for the main claim: under a controlled lightweight YOLO setting, texture residual prior coding is more effective than color-only prior coding and more reliable than naive dual-prior mixing. The relative gains are further visualized in Fig.3.

Table.2. Formal Held-Out Test-Set Comparison for M0-M4

V.	P	R	mAP50	mAP50-95	F1	FPS	Params (M)
M0	0.3812	0.3027	0.2776	0.1341	0.3375	43.45	2.3787
M1	0.3575	0.3159	0.2915	0.1444	0.3354	43.17	2.3787
M2	**0.4851**	**0.3311**	**0.3289**	**0.1542**	**0.3936**	**40.57**	**2.3787**
M3	0.4303	0.2987	0.3105	0.1470	0.3526	37.91	2.3787
M4	0.3883	0.3057	0.2861	0.1399	0.3421	38.02	2.3787

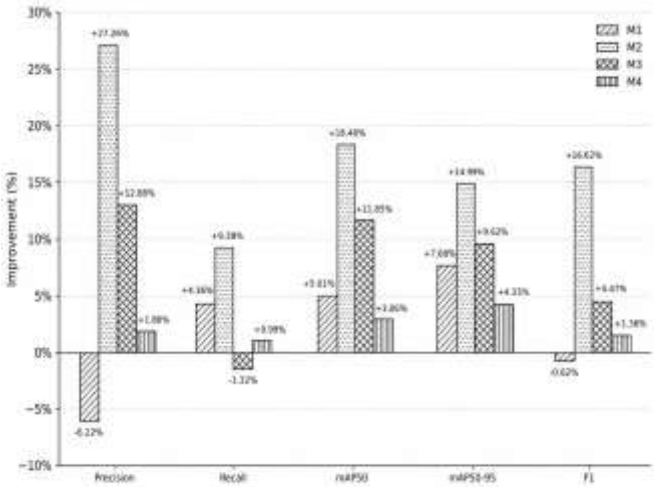

Fig.3. Relative metric gains over the M0 RGB baseline.

M1 improves mAP50 slightly over M0, but the gain is limited and accompanied by lower precision. M3 improves over the baseline but does not surpass M2, indicating that fixed dual-prior mixing does not automatically produce complementary benefit. M4 is only marginally above M0 in mAP50 and remains clearly below M2 and M3, suggesting that the current preprocessing-based prior-adapter route is not the most competitive formulation in the present setting.

5.2 VALIDATION-STAGE CONSISTENCY

Although the final claims in this paper are tied to the held-out test split, the validation-stage ablation trends are also informative because they show that M2 already exhibited the strongest tendency during training-stage inspection.

This consistency between validation behavior and held-out test performance strengthens the interpretation that the anomaly-derived texture prior captures useful structure-sensitive information rather than producing an accidental gain in a single evaluation.

5.3 PER-CLASS OBSERVATIONS

To better understand where the improvement originates, Table 3 compares selected per-class AP50 values for M0 and M2 on the held-out test split. The largest improvements are concentrated in categories that are visually associated with structural disruption, local deformation, or sharply bounded irregular patterns.

Table.3. Selected Per-Class AP50 Comparison Between M0 and M2 on the Held-Out Test Split

Class	M0	M2	Abs.	Rel.
float	0.1713	0.4241	+0.2528	+147.64%
thick_warp	0.1585	0.3808	+0.2223	+140.27%
star_jump	0.0649	0.2438	+0.1789	+275.84%
abrasion_finish	0.2835	0.4379	+0.1544	+54.44%
hundred_feet	0.0990	0.0599	-0.0391	-39.53%
stain	0.2020	0.1054	-0.0966	-47.81%

These gains are consistent with the intended role of the anomaly prior: it amplifies local structural deviation and texture disruption more effectively than the color prior alone. The class star_jump, for instance, benefits by 275.84% relative gain, and float improves by 147.64%. However, the gain is not uniform. Several classes degrade under M2, including stain (-47.81%) and hundred_feet (-39.53%), suggesting that anomaly-only coding is not universally optimal for every defect type, especially when the relevant visual evidence is more diffuse or less aligned with strong structural irregularity. The strongest defensible conclusion is not that the anomaly prior dominates every category, but that it is the best overall prior among the tested variants in this controlled setting.

5.4 QUALITATIVE COMPARISON

Strict same-image qualitative comparison figures were exported on a fixed set of held-out test images. Each panel contains the original image, ground-truth overlay, and predictions from M0-M4 on exactly the same sample (Fig.4). Two useful observations emerge. First, M2 more consistently highlights localized defect regions that are weak or ambiguous under the RGB-only baseline, especially in cases with evident texture disturbance. Second, M1 tends to help less consistently, aligning with its weaker quantitative gain. M3 and M4 remain informative ablation variants, but their improvements are less uniform than those of M2.

Fig.4. Same-image qualitative comparison on a held-out test sample.

In addition to detector outputs, prior-visualization panels were exported showing the original RGB image, the Delta E map, the anomaly map, and the corresponding prior-coded inputs for M1, M2, and M3 (Fig.5). These visualizations help explain why M2 performs best: compared with the Delta E prior, the anomaly prior concentrates more strongly on local structural irregularity and texture disruption, which is consistent with the dominant appearance of many fabric defects in this dataset.

Fig.5. Representative prior maps and prior-coded inputs.

5.5 SPEED-ACCURACY TRADEOFF

The measured FPS values show that M2 improves detection quality with only a moderate reduction in throughput relative to M0. Since all variants share the same YOLO26n detector and parameter count, the speed-accuracy comparison primarily reflects the effect of input coding strategy rather than architectural scaling. M2 therefore remains attractive as a lightweight practical option: it delivers the best held-out test accuracy while preserving the same compact detector size and competitive measured speed.

6. LIMITATIONS AND FUTURE WORK

The current study has several clear boundaries. First, the results come from a single training run per variant under a fixed seed; repeated-seed statistics were not part of the current experimental package. Second, the study focuses on two image-derived priors and does not yet compare them with a wider set of classical texture operators such as Sobel, Laplacian, local binary patterns (LBP) [19], or Gabor-filter features [20]. Third, M4 was evaluated as an offline preprocessing route rather than as a separately validated end-to-end learnable fusion module, so its current result should not be generalized to all learnable prior-fusion approaches. Fourth, the present study focuses on controlled prior-coding ablation rather than detector-family benchmarking; comparisons with other detector families are needed before making broader deployment claims.

These limitations suggest natural directions for future work. A stronger next step would be to compare the anomaly prior with additional classical texture priors, test repeated-seed stability, and investigate adaptive fusion mechanisms that preserve the favorable properties of the current input-level formulation while allowing more selective integration of color and texture cues.

7. CONCLUSION

This paper presented CAP-YOLO, a prior-coded input framework for lightweight YOLO-based fabric defect detection. The study compared five variants under the same YOLO26n detector, data split, and training protocol, with formal claims tied to the held-out test split. Among the evaluated variants, the anomaly-derived texture residual prior (M2) achieved the strongest overall result, improving mAP50 by 18.48% and F1 by 16.62% over the RGB baseline while keeping the detector unchanged at 2.3787M parameters.

The main practical conclusion is that, for the current fabric defect detection setting, a simple image-derived texture residual prior is more useful than a color-only prior and more effective than naive dual-prior mixing. This finding is meaningful for lightweight industrial vision because it improves detection without requiring detector redesign. At the same time, the study also clarifies that preprocessing-based prior fusion has limits, and that stronger adaptive fusion remains open for future investigation.

REFERENCES

[1] Y. Kahraman and A. Durmusoglu, "Deep Learning-Based Fabric Defect Detection: A Review", *Textile Research Journal*, Vol. 93, No. 5-6, pp. 1485-1503, 2023.

[2] C. Li, J. Li, Y. Li, L. He, X. Fu and J. Chen, "Fabric Defect Detection in Textile Manufacturing: A Survey of the State of the Art", *Security and Communication Networks*, Vol. 2021, pp. 1-13, 2021.

[3] H.Y.T. Ngan, G.K.H. Pang and N.H.C. Yung, "Automated Fabric Defect Detection - A Review", Image and Vision Computing, Vol. 29, No. 7, pp. 442-458, 2011.

[4] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han and G. Ding, "YOLOv10: Real-Time End-to-End Object Detection", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1-6, 2024.

[5] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu and J. Chen, "DETRs Beat YOLOs on Real-Time Object Detection", *Proceedings of the IEEE/CVF International Conference on Computer Vision and Pattern Recognition*, pp. 16965-16974, 2024.

[6] Ultralytics, "Ultralytics YOLO Documentation", Available at: https://docs.ultralytics.com, Accessed on 2026.

[7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149, 2017.

[8] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1-5, 2018.

[9] A. Bochkovskiy, C.Y. Wang and H.Y.M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 10934-10945, 2020.

[10] T.Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollar, "Focal Loss for Dense Object Detection", *Proceedings of IEEE International Conference on Computer Vision*, pp. 2999-3007, 2017.

[11] V. Zavrtanik, M. Kristan and D. Skocaj, "DRAEM - A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8330-8339, 2021.

[12] Z. Liu, Y. Zhou, Y. Xu and Z. Wang, "SimpleNet: A Simple Network for Image Anomaly Detection and Localization", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20402-20411, 2023.

[13] J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran and O. Dabeer, "WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19606-19616, 2023.

[14] P. Bergmann, M. Fauser, D. Sattlegger and C. Steger, "MVTec AD - A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9584-9592, 2019.

[15] T. Defard, A. Setkov, A. Loesch and R. Audigier, "PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization", *Proceedings of International Conference on Pattern Recognition Workshops and Challenges*, pp. 475-489, 2021.

[16] K. Roth, L. Pemula, J. Zepeda, B. Scholkopf, T. Brox and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318-14328, 2022.

[17] G. Sharma, W. Wu and E.N. Dalal, "The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations", *Color Research and Application*, Vol. 30, No. 1, pp. 21-30, 2005

[18] Alibaba Cloud Tianchi, "Fabric Defect Detection Dataset", Available at: https://tianchi.aliyun.com/dataset/79336, Accessed on 2026.

[19] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 971-987, 2002.

[20] B.S. Manjunath and W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, pp. 837-842, 1996.