Research Article

Open Access

CrossMark

Reliable fabric defect detection via Bayesian uncertainty modeling

Abstract

Despite the demonstrated capability of deep learning models in detecting anomalies in textile images, their predictions in real-world applications tend to be overly confident, especially when faced with defect types not previously encountered in the training set or when dealing with low-quality annotations. This excessive confidence in predictions limits the practical application of deep learning methods in textile defect detection, as it fails to provide inspectors with reliable guidance on when to trust the model's predictions and when manual verification is necessary. To address this issue, this paper introduces a Bayesian fabric anomaly detection model that utilizes Variational Inference (VI) to apply Bayesian inference to the widely used U-Net architecture. During the inference phase, the model employs Monte Carlo sampling to perform multiple forward passes, generating three types of uncertainty estimations and per-pixel uncertainty maps, thus providing comprehensive evidence for decision-making. This method not only estimates the uncertainty of model predictions but also improves the F1 score by 2-4% over the baseline U-Net model in the frequency domain. This study proves the Bayesian approach boosts fabric anomaly detection and decision-making by optimizing model performance and reducing reliance on inaccurate predictions.

Keywords: Bayesian deep learning, textile anomaly detection, uncertainty estimation, variational inference, U-net architecture

Zhewei Chen,[1] Wai Keung Wong,[2] Jinpiao Liao,[1] Ying Qu[1]

[1]School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong
[2]School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, and Laboratory for Artificial Intelligence in Design, Hong Kong

Correspondence: Calvin Wong, School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, and Laboratory for Artificial Intelligence in Design, Hong Kong, Email calvin.wong@polyu.edu.hk

Abbreviations: VI, variational inference; IOU, intersection over union

Introduction

Anomaly detection in textiles is a critical aspect of fabric quality control, where inspectors typically need to locate and mark defects within rolls of fabric to prevent defect areas from moving on to subsequent cutting and sewing stages.[1] To improve the efficiency and accuracy of defect detection, numerous studies now utilize artificial intelligence and computer vision techniques for the automatic detection.[2] Deep learning methods, with their robust feature extraction and data fitting capabilities, have achieved remarkable accuracy across various fabric defect datasets.[3–5]

Despite the notable success of deep learning methods in the domain of fabric anomaly detection, traditional deep neural networks still face several critical limitations. Firstly, although several defect datasets are accessible to the public,[6–8] their limited size and lack of diversity in defect types and appearances do not fully represent the complexity encountered in real-world applications. This leads to deep learning models often exhibiting overconfident predictions when encountering defect types not seen in the dataset, as well as an inability to accurately identify these unknown defect types.[9,10] Secondly, the performance of deep learning models heavily depends on the quality of data annotation.[11] In the task of fabric defect detection, obtaining precise pixel-wise annotated data is both costly and time-consuming, and inevitable annotation errors directly impact the model's segmentation performance, leading to inappropriate confidence levels in predictions. These issues not only limit the effectiveness of deep learning methods in practical applications but also pose challenges to the automated fabric quality control process. Inappropriate confidence levels in predictions can lead to confusion in decision-making, as fabric inspectors are unable to decide which model predictions can be trusted and which require manual verification.[12] Additionally, inspectors are unable to adjust confidence thresholds to accommodate various inspection standards.

To calibrate the confidence output by models, some studies have proposed generating probability estimates from deep neural networks as measures of model confidence.[12,13] Additionally, popular metrics such as Expected Calibration Error[14] and Maximum Calibration Error[15] can be used to quantitatively assess model calibration. However, these metrics, based on Softmax probabilities, fail to capture epistemic or model uncertainty.[16] To address this challenge, Bayesian deep learning methods have been adopted for effectively capturing uncertainty in image segmentation tasks, notably through Monte Carlo (MC) Dropout[17] to estimate prediction uncertainty. However, concerns have been raised regarding MC Dropout's ability to accurately represent model uncertainty, as it uses dropout to simulate posterior distributions, leading to debates on whether it captures true model uncertainty or just prediction variability due to its inherent randomness.[18] Therefore, considering the common occurrence of small-scale datasets and the challenge of low-quality annotations in fabric anomaly detection and inspired by,[19,20] this study aims to explore a Bayesian deep learning method designed to precisely quantify uncertainty in fabric anomaly detection with minimal effect on model performance.

In this paper, we address the challenge of uncertainty estimation in fabric defect segmentation by introducing a Bayesian fabric anomaly detection model. This model leverages Variational Inference (VI)[21] techniques to enable efficient Bayesian inference within the popular U-Net[22] architecture. During the training phase, VI specifies a parametrized family of distributions and then adjusts these parameters to make one of the distributions in this family as close as possible to the target posterior distribution. In this way, VI transforms the originally complex problem of computing the posterior distribution into a relatively simple optimization problem, making Bayesian inference feasible in high-dimensional spaces and on large datasets. During the inference phase, Monte Carlo sampling[23] is used to draw samples

from the parameter's posterior distribution. Through this process, our model generates multiple predictions per pixel by sampling from the approximate posterior distribution, enabling a direct and quantifiable assessment of uncertainty. The proposed model has been validated on two public fabric defect datasets, with experimental outcomes illustrating its ability to compute three distinct types of uncertainty—MC sample variance, predictive entropy, and mutual information. Moreover, it provides a per-pixel uncertainty estimation, adding depth to our understanding of the model's predictions. Compared to the frequency-domain baseline U-Net model, our approach achieves a significant 2-4% increase in the F1 score. Additionally, this study explored the correlation between segmentation accuracy and the calculated uncertainty estimates, further substantiating the method's robustness and reliability. In summary, this research demonstrates that the proposed Bayesian U-Net can accurately capture the uncertainty in model predictions while ensuring the segmentation performance is maintained.

Methods

Figure 1 illustrates the operational flow of our Bayesian U-Net model for fabric anomaly detection. Beginning with a textile image input, the data is processed through a network of Bayesian convolutional layers that are adept at identifying complex patterns and potential anomalies. Each layer within these Bayesian convolutions employs weights and biases sampled from Gaussian distributions, essential for capturing the uncertainties during the learning process. This structure includes fundamental elements such as skip connections and transposed convolutions, which are crucial for the model's powerful feature extraction and precise segmentation abilities. The multiple sample predictions, depicted on the right, culminate in an uncertainty map that visually conveys the model's varying confidence levels across different segments of the input. The subsequent sections offer a detailed explanation of Bayesian neural networks, Variational Inference, and the three types of uncertainty measurements studied in this research.
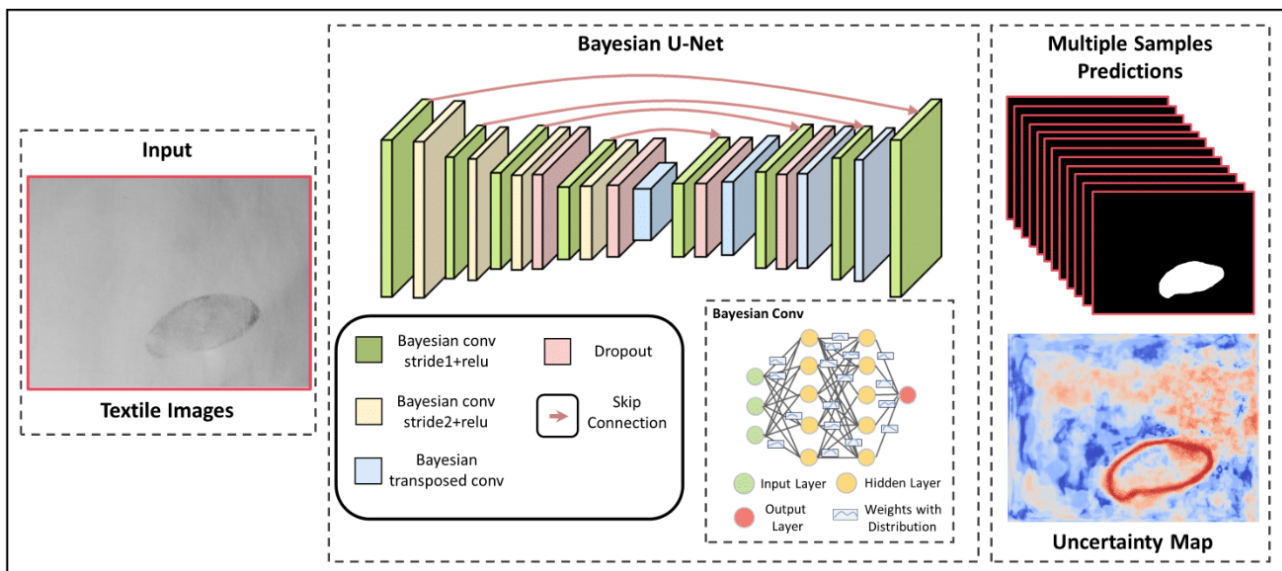

Figure 1 Overview of the proposed Bayesian U-Net for textile defect detection.

Bayesian neural networks

Bayesian Neural Networks offer a probabilistic perspective to deep learning by assigning probability distributions over the weights of a neural network.[23] Given a training dataset $D = \{x, y\}$, where inputs $x = \{x_1, x_2, x_3, x_4, \ldots, x_N\}$ and their corresponding outputs $y = \{y_1, y_2, y_3, y_4, \ldots, y_N\}$. Within a Bayesian framework, the task is to deduce the distribution of weights, denoted as ω, which dictate the function $y = f_\omega(x)$, characterizing the model. Prior to the observation of data, the weights are imbued with a prior distribution $p(\omega)$, reflecting our initial assumptions about the parameters responsible for generating outputs. Armed with the evidence from the data $p(y \mid x)$, along with this prior and the likelihood $p(y \mid x, \omega)$, the objective is to infer the posterior distribution $p(\omega \mid D)$ for the weights. Direct computation of this posterior is usually impractical, necessitating the exploration of alternative inference strategies, such as employing Monte Carlo sampling for approximations. By performing multiple stochastic forward passes and employing Monte Carlo estimators to sample from this posterior distribution of weights,

the predictive distribution can be derived. Given a new input x^*, the predictive distribution of the output y^* is approximated as:

$$p(y^* \mid x^*, D) \approx \frac{1}{T} \sum_{i=1}^{T} p(y^*|x^*, \omega_i), \omega_i \sim p(\omega \mid D) \qquad (1)$$

Here, ω_i represents samples drawn from the posterior distribution $p(\omega \mid D)$, with T denoting the total number of Monte Carlo samples utilized.

Variational inference

Variational Inference is a strategy for simplifying the task of approximating the intricate probability distributions over neural network weights $p(\omega \mid D)$. This method proposes a more tractable distribution $q_\theta(\omega)$, and refines it by minimizing the Kullback-Leibler (KL) divergence from the true posterior. In this process, minimizing the target function effectively means optimizing the Evidence Lower Bound (ELBO), which is defined as:

$$L := \int q_\theta(\omega) \log p(y|x, \omega) d\omega - KL[q_\theta(\omega) \| p(\omega)] \qquad (2)$$

In mean-field variational inference, each weight is represented by an independent Gaussian distribution with its own variational parameters, mean μ, and variance σ^2:

$$q_\theta(\omega) := N(\omega \mid \mu, \sigma^2) \tag{3}$$

The optimization of ELBO, carried out by stochastic gradient descent, enables the learning of both the form of the variational distribution $q_\theta(\omega)$ and its parameters μ and σ, leading to an effective approximation of the model's uncertainty.

Uncertainty measurement in deep learning networks

Three types of uncertainty measurements are computed: MC sample variance, predictive entropy, and mutual information.

MC sample variance: Building on the methodologies established in previous research leveraging Monte Carlo sampling techniques,[24–26] the Monte Carlo sample variance serves as a metric of uncertainty. It is calculated from the variance observed across T Monte Carlo samples from the model's predictive output. The variance for the estimated output $\hat{Y}$ is computed as follows:

$$Var[\hat{Y}] = \frac{1}{T-1} \sum_{t=1}^{T} (p(y^* \mid x^*, \omega_i) - \overline{p(y^* \mid x^*)})^2 \tag{4}$$

Predictive entropy: This metric quantifies the informational content embedded in the model's predictions for each pixel, reflecting the level of certainty it possesses about its estimations. To approximate the entropy for a given pixel x_i^*, the subsequent estimator is employed[27]:

$$H[\hat{y}_l \mid x^*, X, Y] \approx -(\frac{1}{T}\sum_{t=1}^{T}(p(\hat{y}_l = 1 \mid x_i^*, \omega_t)\log(\frac{1}{T}\sum_{t=1}^{T}p(\hat{y}_l=1|x_i^*,\omega_t)) +$$

$$(1 - \frac{1}{T}\sum_{t=1}^{T}(p(\hat{y}_l = 1 \mid x_i^*, \omega_t)\log(1 - \frac{1}{T}\sum_{t=1}^{T}p(\hat{y}_l=1|x_i^*,\omega_t)) \tag{5}$$

This uses the predictive probabilities $p(\hat{y}_l = c \mid x_i^*, \omega_t)$ obtained from the sampled weights ω_t, which in the context of variational inference would be sampled from $q_\theta(\omega)$.

Mutual information: The mutual information represents the shared information between the model's posterior density and its predictive density for every pixel, calculated by taking the difference between the expected predictive entropy and the average entropy of the model's predictions[27]:

$$MI[\hat{y}_l, \omega \mid x^*, X, Y] \approx H[\hat{y}_l \mid x^*, X, Y] - \frac{1}{T}\sum_{t=1}^{T} H[\hat{y}_l \mid x^*, \omega_t] \tag{6}$$

Where $H[\hat{y}_l \mid x^*, \omega_t]$ is the entropy of the predictive distribution for a single sample ω_i from $q_\theta(\omega)$.

Experiments and results

Experiment data

Two publicly available datasets for fabric defect detection were used to evaluate the proposed method: Fabric Stain dataset[7] and the AITEX dataset.[8] Both are publicly accessible and serve as standardized dataset for research.

The Fabric Stain dataset was initially equipped with annotations for defect bounding boxes. Subsequently, pixel-level annotations were produced from these bounding boxes using the LabelMe tool. Experts manually outlined the precise contours within each box, providing detailed pixel-level defect annotations. The AITEX dataset provides pixel-level annotations and encompasses a wide variety of defect types and samples, enhancing its utility for research purposes.

In terms of dataset specifics, the Fabric Stain dataset comprises 394 defect images with corresponding labels, while the AITEX dataset includes 185 labeled defect images. Each dataset is structured into subsets for training (60% of the images), validation (20%), and testing (20%).

Training procedure

In the experiments, the input size for all models was standardized to 512x512 pixels, utilizing the letterbox resize method. Images smaller than this dimension were not scaled up but padded to maintain size consistency.

For model optimization, including the introduced Bayesian U-Net and the comparative baseline U-Net model, the AdamW[28] optimizer was employed. The settings included a learning rate of 1e-3 and beta parameters set to 0.937 and 0.999. Training sessions processed mini-batches of 8 samples each on a 24GB GPU. The limitation of processing capacity on a single GPU necessitated the use of gradient accumulation for batch updates.

This research implemented a cosine annealing with restarts strategy for learning rate adjustment, setting the cycle length at 10 epochs and employing a multiplier of 100. The lowest learning rate was determined to be one percent of the initial rate. Models underwent training for up to 1000 epochs, incorporating an early stopping mechanism to save the checkpoint yielding the highest F1 score on the validation dataset. The U-Net model utilized ResNet101[29] as the backbone network and was initialized with weights pretrained on the ImageNet dataset.[30]

To prevent the risk of overfitting, a series of data augmentation strategies were integrated into the training process. This includes horizontal and vertical image flips, each with a 50% chance, and 90-degree rotations, also at a 50% probability. Adjustments to the image's brightness and contrast were randomly applied, within a variance of 0.2 and a 50% likelihood of being enacted.

Experiment results

Firstly, a comparison was made between the segmentation performance of baseline frequency domain U-Net and the proposed Bayesian U-Net. As shown in Table 1, the Bayesian U-Net model consistently outperforms the standard U-Net in terms of accuracy and F1 score across the Stain and AITEX datasets. Specifically, it shows a 0.4% increase in accuracy and a 2% improvement in F1 score for the Stain dataset, while for AITEX, the gains are 0.5% and 4.2%, respectively. These enhancements are evident in the model's recall and precision; the Bayesian U-Net's higher recall indicates better true positive identification, and its increased precision suggests fewer false positives. The Intersection over Union (IOU) metric also reflects superior performance, with the Bayesian model achieving about a 5% higher IOU on AITEX, signifying greater alignment with the ground truth. From the visual results presented in Figure 3, it is clear that the Bayesian U-Net model effectively identifies and represents the varying levels of uncertainty in both the Stain and AITEX datasets. Notably, the uncertainty maps for the Stain dataset pinpoint regions of higher uncertainty predominantly along the segmentation borders, mirroring the logits variance. For the AITEX dataset, despite the complex patterns of fabric textures, the Bayesian U-Net demonstrates robust segmentation accuracy, evidenced by the detailed uncertainty maps and high F1 scores. In conclusion, the Bayesian U-Net showcases an impressive ability to enhance segmentation performance while

simultaneously providing meaningful uncertainty quantification. The improvements observed across accuracy, precision, recall, and IOU metrics affirm that the Bayesian approach not only refines segmentation quality but also enriches the model's interpretative clarity.

Table 1 Comparative performance of U-net and Bayesian U-net models on stain and AITEX datasets

Dataset	Model	Evaluate metrics				
		Accuracy	Recall	Precision	F1 Score	IOU
Stain	U-Net	0.973	0.809	0.83	0.819	0.694
	Bayesian U-Net	0.977	0.812	0.868	0.839	0.723
AITEX	U-Net	0.963	0.794	0.582	0.671	0.505
	Bayesian U-Net	0.971	0.753	0.677	0.713	0.554

The second part of the experiments delves into the relationship between prediction uncertainty and segmentation accuracy, as quantified by F1 scores. Figure 2 provides a graphical representation of the correlation between logits variance and F1 scores, as measured across two datasets, Stain and AITEX. In both graphs, the data points are scattered, depicting the relationship between the two variables, with the straight line representing the best-fit line derived from linear regression. This line illustrates the trend in the data, showing the direction and strength of the relationship. Quantitatively, the Stain dataset reveals a Pearson correlation coefficient (r) of -0.649, with the near-zero p-value signaling a strong negative correlation, suggesting that a higher logits variance is typically associated with lower F1 scores. A similar pattern is observed in the AITEX dataset, which demonstrates an even stronger negative correlation with an r value of -0.777. The analysis of Figure 3, particularly within the Stain dataset, identifies a pattern indicating that higher segmentation accuracy, as reflected by an increased F1 score, correlates with lower predictive uncertainty. For instance, the second row sample showcases a higher F1 score compared to the first row, indicating more precise segmentation. Concurrently, the sample in the first row exhibits greater values across the three measures of uncertainty—logits variance, output entropy, and mutual information—than the sample in the second row. This pattern of inverse correlation is mirrored in the AITEX dataset samples, where the fourth row indicates better segmentation performance with lower uncertainty than the third row. The observed negative correlation between logits variance and F1 scores across the Stain and AITEX datasets indicates that logits variance may serve as a meaningful indicator of performance in segmentation tasks.
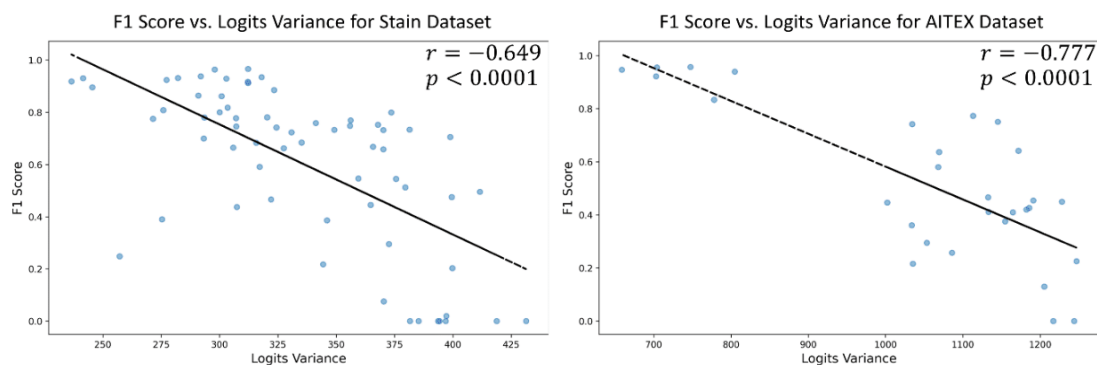

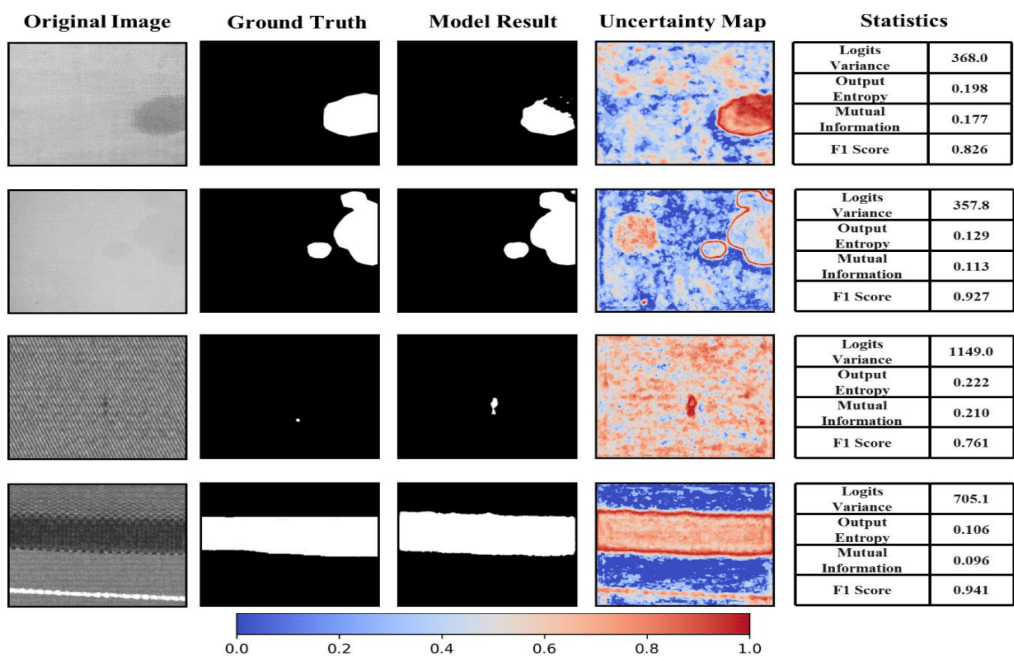**Figure 2** Correlation Analysis of F1 Score and Logits Variance on Stain and AITEX Datasets.

Figure 3 Bayesian model outputs and uncertainty maps for Stain (first two rows) and AITEX (third and fourth rows) datasets.

Discussion

Comparative segmentation performance across datasets

Compared to the frequency domain U-Net model, the Bayesian U-Net exhibits significant performance improvements in fabric defect segmentation tasks. The underlying mechanism for this enhancement lies in the Bayesian model's inference process, which employs multiple Monte Carlo sampling (50 times in this study). This procedure is akin to an ensemble inference from 50 different segmentation models, making the final predictive outcome more robust and reliable. Monte Carlo sampling not only provides a probabilistic prediction but also bolsters the model's generalization capability, as it captures varied model behaviors with each sampling, offering richer information in areas of greater uncertainty.

By utilizing ensemble inference, the Bayesian U-Net more effectively combines multiple predictions, thereby reducing the likelihood of overfitting or bias that might arise in a single model. In the experiments, this approach surpasses the traditional U-Net model in key performance metrics such as accuracy, precision, and F1 score. This advantage is especially pronounced when dealing with images that have complex textures and ambiguous boundaries.

In conclusion, the use of Monte Carlo sampling for ensemble inference of proposed Bayesian U-Net significantly enhances segmentation performance, offering a robust and generalized approach that outperforms traditional frequency domain U-Net models.

Uncertainty estimation and decision-making

Firstly, our results have revealed a negative correlation between segmentation performance and uncertainty. This highlights the utility of uncertainty measures in evaluating model predictions, particularly when ground truth is unavailable. By assessing uncertainty, we can infer the reliability of model predictions, which is especially valuable when the model lacks confidence in its output. In essence, uncertainty serves as an alternative metric to gauge prediction accuracy, providing an evaluative measure in scenarios where direct validation of model results is not feasible. Moreover, consideration of uncertainty enhances model transparency, allowing users to understand and trust the decision-making process of the model.

In practical fabric inspection systems, the application of uncertainty has substantial real-world relevance. Setting a threshold for uncertainty facilitates a straightforward selection mechanism: predictions that exceed a certain level of uncertainty are flagged for review by inspection personnel, while those below the threshold are deemed reliable, thus requiring no further manual intervention. This approach not only improves the efficiency of the inspection system but also ensures that each manual review is value-adding. Importantly, it introduces human intuition and expertise into the AI system's judgments, forging a new model of human-machine collaboration that is particularly beneficial when the model is insufficient to resolve issues on its own.

In conclusion, the significance of uncertainty estimation extends beyond merely enhancing the trustworthiness of predictions. It also provides direction for ongoing improvement of the model, enabling researchers to identify and target areas where the model struggles the most. As this method is adopted in more practical applications, we can anticipate the creation of more intelligent and adaptive machine learning systems. These systems will not only demonstrate resilience in the face of uncertainty but will also foster more meaningful interactions with human users.

Leveraging uncertainty for annotation refinement

The Bayesian model's architecture is inherently designed to resist label noise. This resilience stems from the model's probabilistic nature, where multiple Monte Carlo samples contribute to the final prediction. Such an approach tends to smooth out the effects of incorrectly labeled data, as the influence of any single noisy label is diminished when averaged over many probabilistic predictions.

By leveraging uncertainty metrics and visual uncertainty maps, users can identify potential annotation errors. For instance, in the third row of Figure 3, the Bayesian model highlights areas with high uncertainty, which may correspond to ambiguous or incorrect labels. This feature allows practitioners to pinpoint and revisit uncertain predictions for further verification or correction, thus improving the overall quality of annotations.

In summary, uncertainty estimation serves as a critical tool for enhancing the robustness of segmentation models against label noise and for refining the quality of annotations. The Bayesian model not only provides insights into model performance but also aids in the iterative process of improving training datasets, which is essential for developing more accurate machine learning models.

Limitations and future works

A key limitation of the Bayesian model is its substantial resource consumption and extended inference time due to the computational demands of Monte Carlo sampling. This constraint can be significant, especially when deploying the model in real-time applications or on resource-limited platforms.

For future work, the focus will be on researching more efficient Bayesian models and inference methods. The aim is to reduce computational overhead while retaining the benefits of uncertainty estimation. Optimizing these models for faster performance could potentially expand their applicability to a broader range of practical scenarios, including those requiring real-time analysis.

Conclusion

In this study, we introduced the Bayesian U-Net model and thoroughly validated its efficacy in the task of fabric anomaly detection. The results demonstrate that the Bayesian U-Net not only surpasses the frequency-domain U-Net in key performance indicators such as F1 score and IOU but also provides meaningful estimates of uncertainty. These uncertainty assessments serve as a critical reference for judging the credibility of the model's outputs. In practice, the level of uncertainty can be used to determine whether manual review of the model's predictions is necessary. In summary, the Bayesian model significantly enhances segmentation performance while also supporting the reliability of model predictions and facilitating subsequent manual verification processes.

Acknowledgments

Funding

Conflicts of interest

Authors declare that there is no conflict of interest.

References

1. Abouelela A, Abbas HM, Eldeeb H, et al. Automated vision system for localizing structural defects in textile fabrics. *Pattern Recognition Letters*. 2005;26(10):1435–1443.

2. Li C, Li J, Li Y, et al. Fabric defect detection in textile manufacturing: a survey of the state of the art. *Security and Communication Networks*. 2021;1–13.

3. Ho CC, Chou WC, Su E. Deep convolutional neural network optimization for defect detection in fabric inspection. *Sensors (Basel)*. 2021;21(21):7074.

4. Jing J, Wang Z, Rätsch M, et al. Mobile–Unet: An efficient convolutional neural network for fabric defect detection. *Textile Research Journal*. 2020;92(1–2):30–42.

5. Liu Q, Wang C, Li Y, et al. A fabric defect detection method based on deep learning. *IEEE Access*. 2022;10:4284–4296.

6. Kampouris C, Zafeiriou S, Ghosh A, et al. Fine–grained material classification using micro–geometry and reflectance. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part V 14, Springer; 2016:778–792.

7. Pathirana P. Fabric stain dataset. 2020.

8. Silvestre–Blanes J, Albero–Albero T, Miralles I, et al. A public fabric database for defect detection methods and results. *Autex Research Journal*. 2019;19(4):363–374.

9. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:427–436.

10. Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199, 2013.

11. Nassar J, Pavon–Harr V, Bosch M, et al. Assessing data quality of annotations with Krippendorff alpha for applications in computer vision. arXiv preprint arXiv:1912.10107, 2019.

12. Boldt A, Schiffer A-M, Waszak F, et al. Confidence predictions affect performance confidence and neural preparation in perceptual decision making. *Scientific reports*. 2019;9(1):4031.

13. Subramanya A, Srinivas S, Babu RV. Confidence estimation in deep neural networks via density modelling. arXiv preprint arXiv:1707.07013, 2017.

14. Bohdal O, Yang Y, Hospedales T. Meta–calibration: Learning of model calibration using differentiable expected calibration error. arXiv preprint arXiv:2106.09613, 2021.

15. Nixon J, Dusenberry MW, Zhang L, et al. Measuring calibration in deep learning. CVPR workshops, 2019;2(7).

16. Gal Y. Uncertainty in deep learning. 2016.

17. Mukhoti J, Gal Y. Evaluating bayesian deep learning methods for semantic segmentation. arXiv preprint arXiv:1811.12709, 2018.

18. Alarab I, Prakoonwit S, Nacer MI. Illustrative discussion of mc–dropout in general dataset: uncertainty estimation in bitcoin. *Neural Processing Letters*. 2021;53(2):1001–1011.

19. Blundell C, Cornebise J, Kavukcuoglu K, et al. Weight uncertainty in neural network. International conference on machine learning, PMLR, 2015:1613–1622.

20. Krishnan R, Subedar M, Tickoo O. Specifying weight priors in bayesian deep neural networks with empirical bayes. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(4):4477–4484.

21. Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. *Journal of the American statistical Association*. 2017;112(518):859–877.

22. Ronneberger O, Fischer P, Brox T. U–net: Convolutional networks for biomedical image segmentation. Medical image computing and computer–assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18, Springer; 2015:234–241.

23. Neal RM. Bayesian learning for neural networks. Springer Science & Business Media, 2012.

24. Kendall A, Badrinarayanan V, Cipolla R. Bayesian segnet: Model uncertainty in deep convolutional encoder–decoder architectures for scene understanding. arXiv preprint arXiv:1511.02680, 2015.

25. Leibig C, Allken V, Ayhan MS, et al. Leveraging uncertainty information from deep neural networks for disease detection. *Scientific reports*. 2017;7(1):1–14.

26. Tanno R, Worrall DE, Ghosh A, et al. Bayesian image quality transfer with CNNs: exploring uncertainty in dMRI super–resolution. Medical Image Computing and Computer Assisted Intervention− MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part I 20, Springer; 2017:611–619.

27. Gal Y, Islam R, Ghahramani Z. Deep bayesian active learning with image data. International conference on machine learning, PMLR; 2017:1183–1192.

28. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101, 2017.

29. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:770–778.

30. Deng J, Dong W, Socher R, et al. Imagenet: A large–scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, IEEE; 2009:248–255.