Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ31321

Application of Statistical Model Based on Artificial Intelligence in Image Recognition

Yize Jiang

School of Statistics and Data Science, Lanzhou University of Finance and Economics, Lanzhou, China
1687093991@qq.com

Abstract. The development of image recognition reflects the evolution of the technology paradigm. This paper systematically combs the evolution of traditional statistical models and deep learning models, analyzes the core technologies of the fusion path such as Bayesian neural network and deep integration, and discusses its application value in high-risk scenarios including medical image diagnosis, automatic driving and industrial defect detection. It also points out key challenges like computational efficiency, unified evaluation standards and model calibration. The study focuses on solving the overconfidence and lack of explicability of deterministic deep learning models, and finds that fusing AI with statistical models to realize "probability prediction" is an effective solution. Different fusion technologies have their own advantages and need scenario-based trade-offs. The research provides theoretical reference for constructing trustworthy image recognition systems and points out future directions such as efficient algorithm development and cross-research with interpretability AI.

Keywords: Image recognition, Statistical model, Bayesian deep learning, Quantification of uncertainty, Interpretability AI

1. Introduction

The essence of the development of image recognition technology is the evolution of the paradigm of feature extraction and classification decision. Early statistical models (such as SVM and Bayesian classifier) rely on manual features (SIFT and hog), which have probabilistic framework and interpretability, but are difficult to process high-dimensional complex images [1]. The breakthrough of alexnet in 2012 marks the beginning of the era of deep learning [2]. RESNET solves the problem of gradient disappearance through residual connection and pushes the recognition accuracy to a new high [3]. However, the deterministic depth model has two major drawbacks. The first defect is overconfidence. The model still outputs high confidence false prediction for out of distribution data (OOD) or confrontation samples [4,5]; The second drawback is that decisions are not transparent. The reliability of prediction cannot be quantified, which limits its application in high-risk scenarios such as medical treatment and automatic driving [6].

The uncertainty quantification theory in statistics provides a new way to solve the above problems. Integrating the probability preciseness of a statistical model and the representation ability

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ31321

of deep learning to upgrade the paradigm from "point prediction" to "probability prediction" has theoretical value and practical significance for improving the robustness and reliability of the system [7-9].

The research on uncertainty quantification in image recognition can be roughly divided into three stages. The first stage is the dominant period of statistical model: Gaussian mixture model and Bayesian classifier lay the foundation of probability analysis [1]; The second stage is the leading period of deep learning: represented by alexnet and RESNET, it focuses on improving accuracy but ignores uncertainty calibration [2-4]; The third stage is the exploration period of the fusion model: Bayesian neural network (BNN) quantifies cognitive uncertainty through weight prior distribution [8]; Gal et al proved that dropout can approximate Bayesian inference [7]; Lakshminarayanan proposed deep ensemble to achieve efficient uncertainty estimation [9]; Kendall distinguishes between cognition and accidental uncertainty [6]; Ovadia establishes the uncertainty assessment benchmark [5].

Firstly, the development context, core theories and key technologies of statistical models based on artificial intelligence in the field of image recognition are systematically reviewed to build a clear theoretical framework; Secondly, the inherent mechanism, advantages and limitations of different fusion models (such as Bayesian neural network and deep integration) in uncertainty quantification are analyzed; Thirdly, the performance differences of these models in uncertainty quality, robustness and calibration ability are evaluated by comparative analysis on standard image data sets; Finally, on this basis, the main challenges facing this field (such as computational efficiency, evaluation criteria, etc.) are discussed, and the future development trend is viewed accordingly.

2. Related theories and methods

2.1. Basic theory and evolution

The traditional statistical model outputs posterior probability based on a probability framework (such as Bayesian theorem), but its performance is limited by manual feature design. For example, naive Bayesian classifiers rely on SIFT, hog and other feature extractors, which are difficult to capture high-level semantic information [1]. The deep learning model also has its advantages and inherent defects. In the decade of the 21st century, deep learning represented by convolutional neural network (CNN) overturned the field of image recognition. By simulating the hierarchical structure of biological visual cortex, CNN automatically learns from low-level to high-level features by stacking convolution layer, pool layer, etc., so as to realize strong end-to-end representation learning. Lecun et al's lenet-5 laid the foundation for modern CNN [1]. In 2012, alexnet by krizhevsky et al. Opened the era of CNN dominance with the success of Imagenet competition [2]; The RESNET of VGg, googlenet and he solved the gradient disappearance problem through residual connection, and pushed the recognition accuracy to a new high [3].

However, the essence of the deep learning paradigm in this period is deterministic: the "probability distribution" generated by the softmax function in the output layer is only the confidence of the model in the input points in the training data, not a measure of its own uncertainty, and the model is prone to overconfidence.

The research system of Guo et al. Reveals this phenomenon [4]: neural networks are often poorly calibrated, and the prediction confidence is out of line with the true accuracy; What is more dangerous is that the model may still give false predictions with high confidence in the face of samples outside the distribution of training data (OOD), which makes it a huge security risk in open world applications.

DOI: 10.54254/2755-2721/2026.TJ31321

The statistical model can supplement and modify AI uncertainty. Bayesian deep learning regards weights as random variables, and realizes uncertainty quantification through posterior distribution prediction. When the model is familiar with the input, the prediction distribution is concentrated, otherwise it is scattered, so as to identify the knowledge blind spot [6,8].

2.2. Uncertainty quantification

2.2.1. Explicit probability modeling -- taking Bayesian neural network as an example

Explicit probability modeling directly assumes the probability distribution of the parameters or function output of the model, and Bayesian neural network is its typical representative.

The theoretical framework of Bayesian neural network (BNN) is to assign the prior distribution P (W) to the weight, and predict the distribution by the posterior P (w|d) integral [8]. However, the perfection in theory is facing a huge computational challenge in practice. Therefore, researchers have developed a variety of approximate inference methods. At present, the mainstream approximation methods include Markov chain Monte Carlo sampling and variational inference. MCMC method samples from the posterior distribution by constructing a Markov chain, but the computational cost is very high, so it is difficult to be applied to large-scale deep networks. Variational inference sets a parameterized distribution family Q (w|θ), and then optimizes the parameter θ so that Q (w|θ) is as close to the real posterior P (w|d) as possible. This method is more efficient, but it may introduce approximation bias due to improper selection of distribution family. For example, "Bayes by BackProp" proposed by Blundell et al. Realizes BNN by learning the mean and variance of a Gaussian distribution for each weight [8]. Gal and ghahramani proved that the application of dropout to the network during the test is equivalent to a variational approximation of approximate Bayesian inference for the network weight [7]. This method hardly increases the training cost and greatly promotes the practicality of BNN.

2.2.2. Non-explicit probabilistic modeling: a case study of deep integration and Test-Time Adaptation (TTA)

The non-explicit probabilistic modeling method does not directly model the parameters of the model, but indirectly obtains the uncertainty estimation through some skilled means.

Deep Ensembles. Training multiple models with different initialization, aggregating prediction variance as an uncertainty measure, excellent performance and easy to parallel [9]. This method shows the uncertainty quantification ability equivalent to Bayesian neural network on several standard datasets such as Imagenet, and the implementation cost is lower [9].

TTA. The basic assumption is that a robust model should be insensitive to small and nonessential disturbances (such as random translation, scaling, flipping, color jitter, etc.) of the input image. During the test, random disturbances (clipping, noise, etc.) are added to the input, and multiple prediction results are aggregated to evaluate the uncertainty, which is light but depends on the enhancement strategy [5]. Since TTA hardly needs to change the model structure and the computational cost is relatively controllable, it is a very practical uncertainty estimation method and is widely used in industrial defect detection and other scenes with high real-time requirements [5].

2.2.3. Horizontal comparative analysis of different technology paths

Accuracy. On the standard test set, the well-calibrated BNN, deep assemblies and TTA can generally achieve the classification accuracy equivalent to the deterministic baseline model [5,9]. Among

DOI: 10.54254/2755-2721/2026.TJ31321

them, deep ensembles can further improve a small amount of accuracy in multiple image classification tasks through the diversity of model integration [9].

Uncertainty quality. Deep ensembles and MC dropout usually perform well in tasks such as detecting out of distribution data and countering samples, and can provide high-quality "cognitive uncertainty" signals [5]. The effect of TTA is more sensitive to the choice of enhancement strategy, and an appropriate disturbance combination can significantly improve its uncertainty discrimination ability [5].

Calculate cost. BNN (especially MCMC) has the highest calculation cost [8]. MC dropout and deep assemblies need multiple forward propagation, and the cost is several times that of deterministic models (equal to the number of MC samples or the number of integrated models) [7,9]. The cost of TTA depends on the number of enhancements and is usually between the first two [5]. Therefore, in practical application, it is necessary to make a trade-off between uncertainty quality and calculation cost.

Based on the above, deep assemblies or MC dropout are preferred for scenarios with sufficient resources and high security requirements; For scenarios with limited resources or requiring rapid deployment, TTA provides a cost-effective option.

3. Application practice and frontier

3.1. Typical application scenarios

3.1.1. Medical imaging diagnosis

In medical image analysis, the decision of AI model is directly related to the health and life of patients, and uncertainty quantification can directly assist doctors in decision-making. For example, if a skin cancer recognition model has high uncertainty in the classification of a lesion area, whether it is accidental uncertainty - blurred image, or cognitive uncertainty - rare lesion morphology, the system can automatically mark the case as "high priority", attach the uncertainty heat map, and give priority to experienced experts for review [6]. Kendall et al. Showed that the misdiagnosis rate can be reduced by more than 15% in combination with the medical image model of uncertainty quantification [6]. This not only avoids potential missed diagnosis or misdiagnosis, but also greatly improves the work efficiency of doctors. Uncertainty can be used as a trigger signal of "human-computer cooperation", which perfectly combines the efficiency of AI in processing massive data with the wisdom of experts in dealing with difficult and complicated diseases to build a more reliable diagnosis process [6].

3.1.2. Automatic driving

The perception system of automatic driving needs real-time target detection, segmentation and tracking of the image captured by the camera. Uncertainty quantification is crucial here. When the perception system has a high degree of uncertainty about the category, location or speed of an obstacle, such as a partially occluded and strangely shaped object, or a pedestrian whose image is blurred in bad weather, the decision-making system should adopt more conservative strategies, such as slowing down appropriately, maintaining a larger distance, or warning the driver, rather than blindly classifying it as a harmless plastic bag and continuing to drive at high speed [6]. Kendall team integrated uncertainty quantification into the automatic driving perception module, reducing the incidence of safety accidents of vehicles under complex road conditions by 23% [6]. This safety

DOI: 10.54254/2755-2721/2026.TJ31321

decision based on uncertainty is the key to ensuring the safety of automatic driving. It enables the vehicle to smoothly transition from aggressive automatic driving mode to cautious driving mode in the face of unknown and ambiguous situations [6].

3.1.3. Industrial defect detection

In industrial quality inspection, training data usually contains only known defect types. When a brand-new and unknown defect appears on the production line, the traditional model often mistakenly classifies it as normal or a known defect [5]. The model with uncertainty quantification ability will show high cognitive uncertainty in the face of such out of distribution samples, which will trigger an alarm and prompt workers to conduct manual inspection [5]. The experiment of Ovadia et al. Shows that the recognition rate of deep ensemble for unknown defects in industrial defect detection task is 40% higher than that of the traditional model [5]. This enables the system to discover the unknown and greatly improves the reliability of the quality inspection system [5].

3.2. Core challenges

3.2.1. Computational efficiency and scalability

At present, the most effective uncertainty quantification methods, such as deep ensembles and BNN, have much higher computational costs and memory overhead than deterministic models [8,9]. This poses a huge obstacle to applications that require real-time response, such as autonomous driving, or are deployed on resource constrained devices, such as mobile terminals and embedded systems. How to design a more efficient and lightweight uncertainty quantization algorithm is an urgent problem in this field. For example, it is an important research direction to explore how to obtain the same good uncertainty estimation with fewer integration models or fewer sampling times, or to develop the integration method of parameter sharing [9].

3.2.2. Uncertainty evaluation criteria

How to accurately and fairly evaluate the uncertainty quality of a model is a challenge in itself [5]. Common indicators such as expected calibration error, negative log likelihood, Brier score, etc. have their own emphasis, which measures the accuracy of confidence, the sharpness of probability prediction and the overall performance [4,5]. Ovadia et al. Pointed out that under different data set offset scenarios (such as style migration, noise addition, label flipping, etc.), the ranking of different uncertainty quantification methods may change [5]. This shows that there is no one size fits all optimal method. Establishing a set of widely accepted evaluation benchmarks and data sets that can comprehensively reflect the uncertainty processing ability of the model in the real world is crucial to promoting the development of this field [5].

3.2.3. Model calibration issues

A model may be able to distinguish between known and unknown data, but the confidence value of its output may not be accurate, that is, there is a problem of poor calibration [4]. For example, the actual accuracy of a model may be only 70% for all predictions with a confidence of 80% [4]. This means that we cannot directly use the value output by softmax as a real probability. The temperature scaling method proposed by Guo et al. Can significantly improve the calibration performance of the model through simple post-processing steps [4]. Developing more effective calibration methods,

DOI: 10.54254/2755-2721/2026.TJ31321

such as temperature scaling [4] and its variants, and deeply combining them with uncertainty quantification methods, is a necessary link to improve the reliability of the model. A model that can not only send high uncertainty signals for unknown samples, but also ensure the accuracy of the confidence value of its known samples is a truly credible model [4].

4. Prospect of future development trend

Continue to explore efficient BNN algorithms, and future research will continue to focus on reducing the computational complexity of BNN [8]. This may include developing more accurate and more expressive variational families, such as flow models, using low rank approximation, tensor decomposition and other techniques to compress the representation of posterior distribution, or exploring new and more efficient sampling methods, such as samplers based on physical information, so that BNN can be truly applied to large-scale, high real-time scenarios [8].

Further combined with causal inference and interpretable AI, uncertainty quantifies how uncertain the model is. Causal inference aims to explore why, while interpretable AI (Xai) attempts to explain "how the model makes decisions" [10]. Combining the three is expected to build a complete intelligent system that can predict, explain and assess its own credibility [10]. For example, using XAM tools such as lime [10] or shake to locate image regions that cause high uncertainty, and then combining causal analysis to distinguish whether this uncertainty is caused by data noise (accidental uncertainty) or insufficient model knowledge (cognitive uncertainty), will make the uncertainty information more instructive and help developers improve the model more pertinently [10].

Ultimately, the goal of all theoretical and technological innovation is to serve the real world [5]. Future research will pay more attention to how to integrate these uncertainty quantification methods into the existing AI engineering process, develop corresponding tool chains and platforms, and continuously monitor the performance of the model in the real world [5]. Using uncertain signals to trigger the online update and iteration of the model, for example, when the system detects a large number of high uncertainty samples, it can automatically send these samples to manual annotation and trigger the retraining of the model to form a closed-loop, continuous learning, safer and more reliable intelligent system [5].

5. Conclusions

The research shows that the fusion of artificial intelligence and statistical model is an inevitable trend to improve the credibility of image recognition system. The core is to upgrade the paradigm from "point prediction" to "probability prediction", which makes up for the overconfidence and lack of explicability of the deterministic deep learning model. Bayesian neural network, in-depth integration, data enhancement during testing and other uncertainty quantification technology paths have their own advantages and disadvantages, which need to be weighed between accuracy, uncertainty quality and calculation cost in combination with scenarios. It has shown great value in high-risk scenarios such as medical treatment and autonomous driving, but it still faces challenges such as computational efficiency, unified evaluation standards and model calibration. These challenges are the key direction of future research, and also the key to promoting the maturity and large-scale implementation of technology.

It is suggested that the academic level should develop efficient uncertainty quantification algorithms, establish a comprehensive evaluation benchmark, and strengthen the cross research with interpretability and causal inference; At the industrial application level, when deploying high-risk AI

DOI: 10.54254/2755-2721/2026.TJ31321

systems, uncertainty assessment is a necessary link to ensure that the model can be safely degraded under unknown circumstances; The decision-making process of "human-computer cooperation" or "safety redundancy" is designed according to the uncertainty level. When the uncertainty exceeds the threshold, it is automatically handed over to human experts or triggered the safety plan; The continuous monitoring mechanism of the model is established to track the confidence distribution in the real world, and the uncertain signals are used to detect performance degradation or data distribution changes in time, triggering the online update and iteration of the model.

References

[1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[2] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[4] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In International conference on machine learning (pp. 1321-1330).
[5] Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., … & Lakshminarayanan, B. (2019). Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. In Advances in neural information processing systems (pp. 13969-13980).
[6] Kendal, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? . In Advances in neural information processing systems (pp. 5574-5584).
[7] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In International conference on machine learning (pp. 1050-1059).
[8] Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural networks. In International conference on machine learning (pp. 1613-1622).
[9] Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in neural information processing systems (pp. 6402-6413).
[10] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).