Journal of Big Data

RESEARCH Open Access

Ramifications of incorrect image segmentations; emphasizing on the potential effects on deep learning methods failure

Hayat Al-Dmour*

*Correspondence:
HDmour@mutah.edu.jo

Faculty of Information Technology, Mutah University, Mu'tah, Karak, Jordan

Abstract

Introduction: Detecting failure cases is critical to ensure a secure self-driving system. Any flaw in the system directly results in an accident. In genuine class, the model's probability reflects better-reflected model confidence. As a result, the confidence distributions of failed predictions were changed to lower values. In contrast, accurate predictions were remained associated with high values, allowing for considerably more excellent separability between such prediction types. The study investigates the association of ramifications with computational color constancy that can negatively influence CNN's image classification and semantic segmentation.

Methodology: Image datasets were used to conduct different scales and complexity experiments. For instance, minimal and straightforward images of digits were comparatively provided through MNIST and SVHN datasets. The dataset's standard validation set was employed to test and compute additional metrics because ground truth that is not publicly available for some test sets.

Results: The results depicted that baseline methods were outperformed through the proposed approach with a considerable variant on minimal datasets or models in every context. Therefore, Transmission Control Protocol (TCP) is appropriate in failure prediction, and ConfidNet is competent to be fulfilled as confidence criterion. Further, one of the solutions would be to elevate the validation set size, but this would influence the prediction performance of a failure model. On the contrary, the confidence estimation was based on models with test predictive performance levels, similar to baselines.

Conclusions: The gap between validation accuracy and training accuracy was significant on CIFAR-100, which indicates the modest enhancement for failure detection via the validation set.

Keywords: Deep learning model, Failure detection, Image classification, Image segmentation, Neural network, Ramifications

Introduction

Neural networks are progressively deployed where the misclassification cost is higher when obtaining higher accuracy across various issues, including medical decision-making systems [1]. On the contrary, there is an authentic emphasis on predictability and trust in the test time performance for neural networks to achieve widespread popularity in these domains. Most high-risk applications have developed legacy procedures that can perform the task, such as human professionals making a classification [2]. A critical element to maintain trust in a model's performance is developing estimates in the prediction confidence that emphasize the accurately anticipated accuracy of that sample [3]. This would facilitate a practitioner to not better comprehend the opportunity of the model forecasting incorrectly on a per-sample basis but also likely utilize that estimate for determining when to default to the legacy procedure. There are two core uses for estimating the prediction confidence [4]. Some applications require the confidence estimate directly as a model output, which is utilized in the next phase of the decision-making procedure [5]. Such applications need to represent the expected sample accuracy through confidence estimate and confirm the probability's natural interpretation [6].

Deep neural networks have observed a greater acceptance, led by their significant performance in different tasks such as object recognition, natural language processing, speech recognition, and image classification [7–11]. Despite their growing success, safety is a significant issue in integrating such models in real-world circumstances [11]. Estimating a model error in applications where failure leads to extreme repercussions becomes more crucial, including nuclear power plant monitoring, medical diagnosis, or autonomous driving [12]. In this regard, failure prediction was addressed with deep neural networks [13–15]. From a classification viewpoint, a widely used benchmark had taken the value of the forecasted class's probability, such as the Maximum Class Probability (MCP). MCP for failure prediction still experiences different conceptual limitations even though recent assessments indicate significant performances with deep models [16].

Indeed, SoftMax probabilities are classified as non-calibrated, inadequate to detect distribution examples, and sensitive to adversarial attacks [17]. Another critical concern associated with MCP is based on confidence scores ranking, which is unrealistic for the failure prediction task [18]. The issue must arise because MCP drives by designing toward high confidence values, even for flawed ones. However, the likelihood of the model shows a better-reflected model confidence in terms of true class. This drives to fails' confidence distributions transformed to lesser values, whereas accurate predictions were still related with high values, which allows a much better separability between such prediction types. Therefore, this paper presents a failure prediction model with deep neural networks by introducing a new confidence criterion based on using the Transmission Control Protocol (TCP) in terms of failure prediction to offer theoretical confirmations.

Related work

Learning model confidence for failure prediction

Deep neural networks were used to define appropriate confidence criteria for predicting failed cases, specifically in the case of classification. Semantic image segmentation was further considered, which can be observed as a pixel-wise classification issue, where a dense segmentation mask was reported through a predicted class model allocated to each pixel. In particular, all the following material was developed for classification, and integration details were provided where needed.

Consider a dataset $\mathcal{D}$ comprising $Ni.i.d$ training samples $\mathcal{D} = \left\{\left(x_i, y_i\right)\right\}^{N_{i=1}}$ where $x_i \in \mathbb{R}^d$ is a d-dimensional characteristic and $y_i^* \in \gamma = \{1, \ldots, K\}$ is its actual class. A classification neural network was viewed as a probabilistic model undertaking an input x, parameters of the network w, and the network allocating a probabilistic predictive distribution for each class k. The model can predict the class as $\widehat{y} = argmaxp(Y = k|w, x)$.

Network parameters were obtained following a maximum likelihood estimation model throughout training, where one reduces the Kullback-Leibler (KL) divergence between the actual and predictive distribution. This is comparative to minimize the cross-entropy loss, in classification, concerning w, which is the negative sum of the log-probabilities over positive labels:

$$\mathcal{L}_{CE}(w; \mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} y_i^* logP(Y = y_i^*|w, x_i)$$

Using confidence estimates

Well-calibrated confidence estimates become progressively important since deep learning models get integrated with real-world decision-making systems where the cost of misclassification is high. A confidence estimate is well-calibrated if a sufficiently closer estimate to the probability of that input being accurately classified. For accurate classification, a probability estimate was obtained by obtaining the sample average preciseness of all data points with similar attributes. A grouping can be done on similar inputs in circumstances where there are few data points with similar characteristics. The confidence estimate uses particular applications from a discriminative model as an input to the next phase of the decision-making process.

By learning mapping to a well-calibrated probability from prediction scores. T-scaling, short for temperature scaling, is a specific example of Platt calibration in which the logit score of a classifier is divided by a scalar T. [12] discovered T-scaling to be the most successful and most straightforward calibration approach in a thorough examination of calibration methods. Because T-scaling does not influence prediction rank-order, it only affects the Brier error, anticipated calibration error, instead of the label error. Calibration parameters are fitted to the validation set, identical to the training set. Calibration does not directly address unfamiliar samples, but our studies indicate that calibration is critical for providing appropriate confidence estimates on both known and unfamiliar data.

Jiang et al. [19] determines the continuity of various types of therapy using ICU mortality calculators for confidence estimates. It becomes essential to obtain the similar intuitive meaning individuals would anticipate because the next step is often determined on that assumption. The overall estimated probability distribution can be utilized as an input for another model across all possible classes instead of comparing the confidence estimate of the predicted class to a threshold. An interpretable probability estimate is needed if a human expert recommends that value. However, confidence estimates can be used to determine whether for trusting the predictions of a model in link with a threshold. This can effortlessly be utilized in the example of automated medical diagnoses since the model can depend on a professional for inputs that cannot be estimated with adequate confidence [19]. The model should merely be used when the user can trust the accuracy of its prediction since the legacy process can be treated as the expert's prediction receiving a diagnosis from a doctor, and there might be a high cost for imprecise diagnoses. -In this regard, the confidence estimate doesn't need to be interpretable as an autonomous quantity. Still, it can be utilized to develop a better predictor of trust while predicting the model.

Calibration of modern neural networks

A natural probability distribution is received by applying a SoftMax layer on the neural network's output for classification problems. On the contrary, recent work has indicated that modern neural networks are adversely calibrated despite higher generalization estimates [17]. Several changes were studied to neural network design and training recently and consequently associate this pattern with increases in model capacity and a type of overfitting. A certain increase was noted in the negative log-likelihood (NLL) after a specific point, indicating that the model exceeds the NLL loss irrespective of test accuracy overfitting [17]. This is particularly possible with the NLL loss. The loss can also be reduced by pushing the anticipated probability distribution across output classes even after the correctly classified train points. In particular, the probability anticipates from modern neural networks can be overconfident. These findings are supported with [20], which indicates that deep neural networks can witness the conventionally reinforced idea that large models are poorly generalized irrespective of regularization. Guo et al. [17] have recommended that the overfitting observed during training does not show in the generalization error but rather in the accuracy of confidence estimates. Previous studies have explored confidence estimate calibration through neural networks but need an ensemble model for the objective of calibration becomes expensive [21].

Image improvement techniques

Different techniques can be adopted to improve image quality, such as adjusting contrast and brightness, dodging, and burning (adjusting the brightness in an area), color balance, and cropping [22]. These methodologies are considered traditional techniques. The contrast, colors, and brightness depend on the scene's characteristics, the settings of the devices, and the quality of the components [23]. The non-traditional image enhancement techniques are: filtering linear (linear filtering), non-linear contrast adjustments (non-linear contrast adjustments), random-noise reduction, filter models for noise reduction (pattern noise reduction filters), and color processing [24]. Linear filtering techniques, such as sharpening, deblurring (anti-blur), edge enhancement, and deconvolution (correction technique based on an algorithm that allows reconstruction of the missing elements on a statistical basis, remove the disturbing factors and make it possible to create a higher quality image), they are used to increase the contrast of small details in an image [25].

Non-linear contrast adjustment techniques include gamma corrections, scale transformations of gray, and curves and lookup tables. These techniques are used to adjust the contrast in selected brightness ranges in an image [26]. Random-noise reduction techniques include low pass filters, blurring filters, median, and speckling (creating images from spots). Instead, the patterns of filters for noise reduction (Pattern noise reduction filters) identify patterns that replicate in the image and allow users to remove them selectively. Color processing includes transformations of the color space, pseudo-coloring (pseudo coloring, also called color level coding) of hue, and finally, the adjustment of saturation [27]. These techniques can change the characteristics of objects in an image.

Other approaches further determined the concern of MCP about high confidence predictions in tasks closely associated with failure prediction [17, 21]. Previously, the temperature scaling method was used for mitigating confidence values for out-of-distribution detection and confidence calibration. On the contrary, this does not influence the confidence score ranking and; thereby, the variance between correct predictions and errors exists. A similar objective of learning confidence in neural networks was presented by [4]. The work varies by mutually emphasizing out-of-distribution detection and learning classification probabilities and distribution confidence scores.

Additionally, the predicted confidence score was used for interpolating output probabilities and target, while TCP was defined as an appropriate metric for failure prediction. An adjunct to the Bayesian neural network was proposed by [21] by allowing neural networks to produce well-calibrated uncertainty measures. A proper scoring rule was used as a training criterion for corresponding to a model prediction's exponential cross-entropy loss value.

Many tools are used to enhance images, and these tools are further divided into two techniques: Point Technique and Spatial Technique. The method, called point, has some methods such as contrast, stretching, clipping the noise, modification, and coloring it, which is called pseudo [28]. Most of the time-image processing is used, which is also used in many operations. Another Spatial technique is also used in processing the image. All of the operations used in this technique are called linear operations, which are mainly used today [29]. The main reason for using this technique is that these operations are very easy and straightforward. Their implementation is also not too complex compared to non-linear operations used in the point technique [30]. Non-linear methods are used primarily at the edges of images and to find the complete details, but linear techniques are mainly used to blur and distort. Also, non-linear methods cannot remove noise from those images because they always contain noise due to their randomness [31]. For instance, in the past, many people used images to capture films with some voice which can cause recording the noise, and this noise needed to be removed. When images' signals are generated, the digitization process is also used, which mostly captures the noise [32].

Digital images produce large amounts of data to be stored. Therefore image compression techniques reduce memory requirements by limiting the data to be recorded. Lossless compression (without loss of information) minimizes the size file eliminating redundant information [33]. Therefore, the content of an image is not altered when it is decompressed. Lossy compression (with data loss) achieves a more significant reduction in file size by removing both redundant and irrelevant information. Since the irrelevant ones cannot be reconstructed when viewing an image, this type of compression causes an inevitable loss of image content and the introduction of artifacts [34]. The higher the compression rates, the greater the loss of information.

The objective of the reconstruction is to eliminate a sort of interference present in the image, called noise, understood as the superposition of unwanted signals on the signal of interest [35]. In the presence of noise, the image typically has a grainy appearance. Still, it may contain real "gaps" in the case of salt and pepper noise in which, randomly, a percentage of information in the image is completely lost [36]. Typically, this is caused by problems in signal transmission (as in the case of medical images) or by poor lighting in the scene. The purpose of denoising is to remove interference in the signal, resulting in a defined, noise-cleaned version of the image. All fundamental structures have been maintained, and the noise eliminated [37].

The human body is a complex system, and data acquisition on its static and dynamic properties produces large amounts of information. One of the biggest challenges is acquiring, processing, and displaying data about the body so that that information can be viewed, interpreted, and used to allow its analysis in diagnostic procedures and assist in therapies [38]. In many cases, the presentation of information about the human body in images is the most efficient approach to address this challenge. Medical images are produced by the interaction of some kind of energy with the human body's tissues, organs, or systems [39]. Producing medical images is always related to specific power (electromagnetic, mechanical) interaction with the matter. The image is visualized using a contrast parameter, determined by some physical characteristic that differentiates the different tissues, organs, or systems [40]. Except for ultrasound, which uses mechanical energy, most images interact with electromagnetic energy and the human body.

Study gap

Therefore, this paper presents a failure prediction model with deep neural networks. A new confidence criterion was introduced based on using the TCP for offering theoretical confirmations in terms of failure prediction. A new method was introduced for learning a predefined target confidence criterion from data as the true class was unidentified at test time. Bayesian deep learning and collaborative approaches discussed connections and differences associated with failure prediction work.

The study is significant as it proposes a specific method for learning failure prediction models with deep neural networks with a confidence neural network based on a classification model. The experimental results validate substantial enhancement from strong benchmarks on different semantic and classification segmentation datasets considering the efficacy of the proposed approach.

Methodology

Figure 1 shows the approach presented in this section was assessed for predicting failure in image segmentation and classification. Initially, comparative experiments were performed alongside Bayesian uncertainty estimation and state-of-the-art confidence estimation methods on different datasets. These findings were then conducted by a comprehensive investigation of the impact of the confidence criterion, learning scheme, and training loss in this approach. Lastly, a few portrayals were provided for obtaining further insight into the behavioral approach.

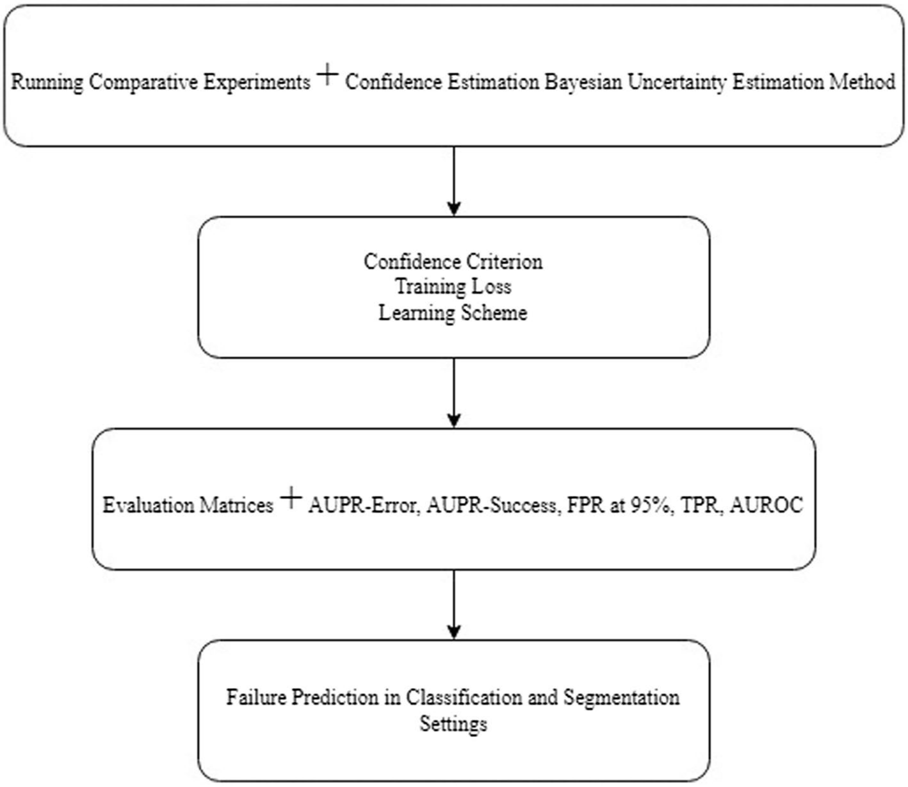
Running Comparative Experiments + Confidence Estimation Bayesian Uncertainty Estimation Method
Confidence Criterion
Training Loss
Learning Scheme
Evaluation Matrices + AUPR-Error, AUPR-Success, FPR at 95%, TPR, AUROC
Failure Prediction in Classification and Segmentation Settings

Fig. 1 Schematic diagram

Experimental data

Data sets

Image datasets were used to conduct experiments on different scales and complexity. For instance, minimal and straightforward images of digits were comparatively provided through MNIST and SVHN datasets [41, 42]. Similarly, additional details were presented regarding object recognition tasks on low-resolution images through CIFAR-10 and CIFAR-100 [43]. Moreover, CamVid [44] was used to report semantic segmentation experiments using a contemporary road scene dataset. The study employed the dataset's standard validation set for testing in some circumstances to compute additional metrics because ground truth is not publicly available for some test sets.

Network architectures

This study has followed the classification of deep architectures as presented by [45] for an appropriate comparison. They vary from minimal convolutional networks for SVHN and MNIST to greater VGG-16 architecture for the CIFAR datasets. The study conducts an investigation to several design architectures of the MLP neural network, which relates to different quality results. Such that, a multi-layer perceptron (MLP) was added with one hidden layer for MNIST to investigate small models' performances. The proposed design structure is likely to be expandable to different hardware specifications and accuracy constraints. Therefore, a SegNet semantic segmentation model was applied for CamVid based on the proposition of [46]. The penultimate classification network layer was connected with ConfidNet, the prediction network model integrated into this study. It is comprised of a succession of five dense layers. Such architecture variants have been investigated, which lead to similar performances. ConfidNet layers were trained before fine-tune the duplicate ConvNet encoder committed for estimating confidence following the learning scheme. ConfidNet was adapted for semantic segmentation by preparing it entirely convolutional.

Assessment parameters

The evaluation of failure prediction was done through predefined parameters as proposed in [17]: AUROC, FPR at 95%, AUPR-Error, and AUPR-Success. In this regard, the core emphasis will be shifted toward AUPR-Error for computing the area under the Precision-Recall curve through the positive class errors.

Results

Comparative findings on failure prediction

Uncertainty estimation and competitive confidence approaches were encompassed for demonstrating the method's effectiveness. These approaches encompass Monte-Carlo Dropout (MCDropout) [10], Maximum Class Probability (MCP) [17], Trust Score [20]. Table 1 summarizes comparative results. Initially, it was observed that baseline methods were outperformed through the proposed approach in every context with a considerable variant on minimal datasets or models. This shows the adequacy and appropriateness of TCP in failure prediction, and ConfidNet is competent to be fulfilled as confidence criterion.

Better results were also presented on minimal datasets or models through the TrustScore method, including MNIST, enhanced baseline. On the contrary, the effectiveness of ConfidNet was majorly seen on larger and complicated datasets, whereas the performance declines for TrustScore due to high dimensionality issues with distances. The number of training neighbors and test samples was drastically reduced through computational complexity, where each training pixel was a neighbor in semantic segmentation.

Random samples were conducted in each train and test image classification for computing a minimal percentage of pixels in TrustScore. On the contrary, ConfidNet showed efficacy in its durability, speed, and output. State-of-art performances were further enhanced, considering confidence measures on dropout layers. Figure 2

Table 1 Comparison of Failure Prediction Model using Different Datasets

Dataset	AUPR-Success	AUPR-Error	AUC	FPR-95%-TPR
MNIST MLP	36.62	35.33	98.04	21.60
MNIST Small ConvNet	43.26	31.03	93.44	75.63
SVHN	63.89	30.04	93.40	58.92
CIFAR-10	151.22	32.05	86.77	51.87
CIFAR-100	90.64	43.51	89.19	38.73
ConfidNet	28.36	61.54	86.84	76.21

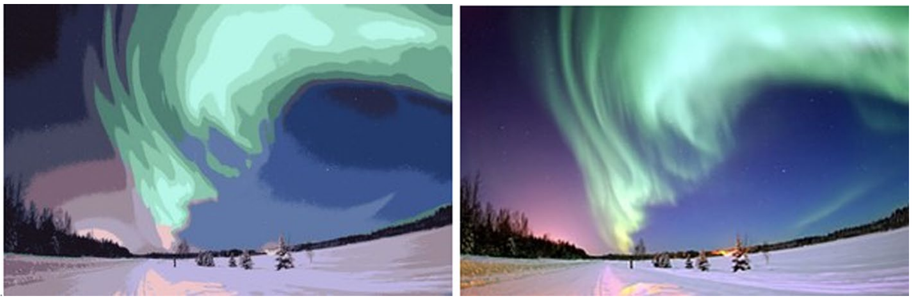

Fig. 2 Large image performances

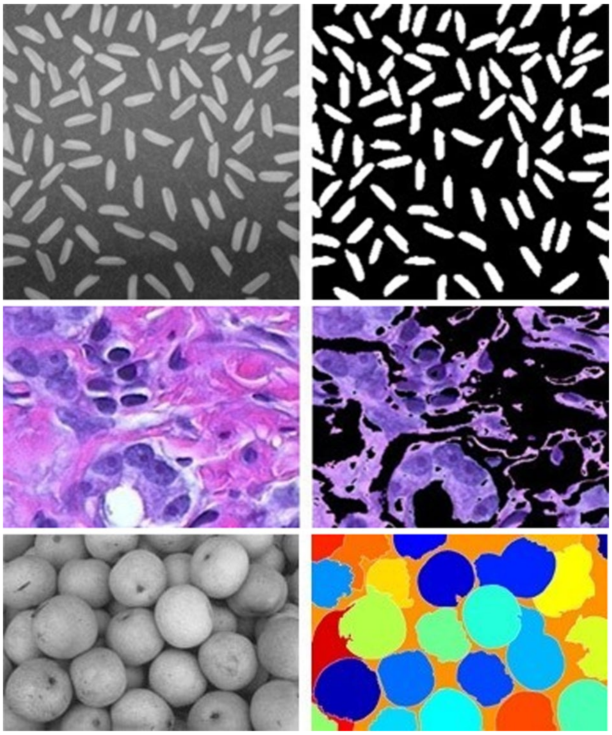

Fig. 3 Texture filters for SHVN and CIFAR10 datasets

showed side-by-side two samples with a similar distribution entropy. A misclassified clustering sample is presented in the left image, whereas an accurate prediction can be shown in the right one. The image confidence was represented from a correct prediction with [0.60, 0.40] distribution, while false confidence was presented with one [0.40, 0.60] distribution. Based on this discussion, an incorrect image can be differentiated from an accurate prediction despite having similar clustering distributions.

Figure 3 portrays the performance of ConfidNet and other metrics for SHVN and CIFAR10 datasets as depicted in risk-coverage curves [8, 11]. A threshold was used as a selection function for corresponding the probability mass of the non-rejected region. From the performance of both datasets, a better coverage was presented by

Table 2 Learning Scheme Effects

	CIFAR-100	MNIST
Confidence training	67.86%	43.49%
Detection performance	68.12%	44.98%

Table 3 ConfidNet comparison on validation and training dataset

AUPR-Error (%)	MNIST (MLP)	MNIST SmallConvNet	SVHN	CIFAR-10	CIFAR-100	CamVid
ConfidNet (Train set)	52.25	48.20	41.32	45.80	35.33	80.95
ConfidNet (validation set)	50.58	59.34	77.15	90.09	31.03	63.00

both datasets for each selective risk that can be selected beforehand by a user. Additionally, the improvement was more pronounced at high coverage rates such as SHVN [0.80, 0.90] and CIFAR-10 [0.7, 0.85] for emphasizing the potential of ConfidNet in identifying critical failures successfully.

Effect of learning variants

The impact of fine-tuned ConvNet was assessed initially in this study. Significant enhancements were fulfilled regardless of fine-tuning in terms of baseline, as presented in Table 2. The performance of ConfidNet was improved in every context by almost 2% after allowing corresponding fine-tuning. It was noted that no significant improvement was brought under consideration regardless of deactivating dropout layers. Training ConfidNet was experimented on a hold-out dataset undertaking the small number of errors available because of deep neural network over-fitting.

Table 3 presents findings for all datasets based on validation sets with 15% of samples. A reduction was observed in a general performance when utilizing a validation set for training TCP confidence. The decline was particularly pronounced for small datasets, where models achieve validated and $\geq 95\%$ trained accuracies. In particular, no more significant absolute number of errors was obtained with a minor validation set and a high accuracy for the validation set compared to the train set. On the contrary, the confidence estimation was based on models with test predictive performance levels, similar to baselines. The gap between validation accuracy and training accuracy was significant on CIFAR-100, which indicates the modest enhancement for failure detection via the validation set. One of the solutions would be to elevate the validation set size, but this would influence the prediction performance of a failure model. It was observed that the approach could be improved by training ConfidNet on the validation set with models reporting low or middle test accuracies.

ConfidNet was trained and then compared with MSE loss to binary classification cross-entropy loss. It was observed that lower performances were accomplished on CIFAR-10 and CamVid datasets, although BCE mainly addresses the failure prediction task. Similar outcomes were also tested and presented through focal loss and ranking loss. It was intuitively observed that training was regularized in TCP by offering additional fine-grained evidence regarding the classifier quality about a sample's prediction.

This was particularly essential in the complex learning configuration where very few error samples were available because of better classifier performance. The effect of regression was further assessed to the normalized criterion. The finding shows the difficulty of correct or incorrect classification training since T CPr was lower than the TCP for small datasets, including CIFAR-10.

Qualitative assessments

A portrayal is represented on CamVid to understand the approach for failure prediction better. Higher confidence scores were produced in this approach for accurate pixel predictions and lower ones on mistakenly forecasted pixels, allowing the user to detect errors effectively in semantic segmentation (Fig. 4).

Discussion

According to the experimental results, the level set approach may obtain an accurate segmentation result with adequate information. When the scene in the photo is more complicated, however, the level set approach cannot produce the necessary segmentation result. As a result, specific pattern recognition algorithms are developed to provide additional information about the target. For example, the target's areas and each pixel's likelihood corresponds to the target category. The level set approach depends on statistics of pixels within and outside the contour throughout the contour evolution process, such as the mean, weighted mean, and probability model regarding areas [47]. The level set approach is more akin to an information integration method in that it employs the energy functional minimization principle to generate a potential function. Furthermore, this possible function can be designated as the probability function, determined using probabilities and Bayesian approaches.

Even though the pixels are relatively similar, the inaccurate probability map and the adjusted prior shape significantly influence, leading comparable pixels to be separated. However, we may use the concept of superpixels [48]. The majority of image segmentation approaches may be summarized as extracting and using information from pictures. As a result, the most significant challenge is to create a dynamic hierarchical organized picture representation.

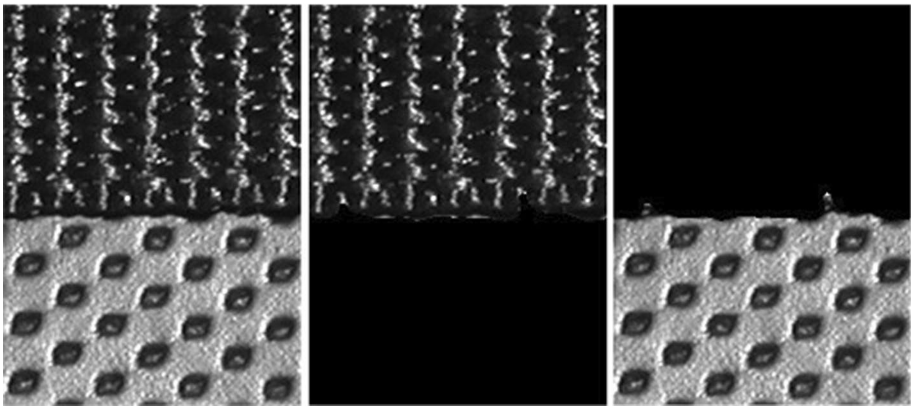

Fig. 4 Inverse confidence patterns

The explicit representation of spatial changes and picture noise in our mathematical formulation of test-time augmentation is based on an image acquisition model. It may, however, be simply adapted to accommodate more generic transformations such as elastic deformations [49] or to include a simulated bias field. In addition to the range of possible model parameter values, the prediction result is also affected by the input data, such as picture noise and object modifications. As a result, a proper uncertainty assessment should consider these elements. For regression problems when the outputs are not discretized category labels, the variance of the output distribution may be more appropriate for estimating uncertainty than entropy.

Conclusions

A new confidence criterion was proposed for the failure prediction model with deep neural networks to offer both empirical pieces of evidence and theoretical guarantees for addressing failure prediction. A specific method was presented with a confidence neural network and application of ConfidNet based on a classification model for learning this criterion. Findings indicated a substantial enhancement from strong benchmarks on different semantic and classification segmentation datasets for validating the efficacy of the proposed approach. The application of ConfidNet can be integrated for estimating uncertainties in multi-task learning and domain adaptation. The majority of image segmentation approaches may be summarized as extracting and using information from pictures. As a result, the most significant challenge is to create a dynamic hierarchical organized picture representation. Furthermore, building a multi-objective matching approach would allow the proposed system to handle more complicated situations.

Additional work is required to refine the offered approach and implement the supplied prototype in the actual circumstance of segmenting the brain tumor. To begin with, just the grey level is used as the deep network's input in this research; in the future, we may use other features, such as texture features, as the deep network's input. Furthermore, additional brain tumor MRI data must be obtained on an ongoing basis. More data will help our suggested technique and other tumor classification systems.

Abbreviations

CNN: Convolutional neural network; MNIST: Modified National Institute of Standards And Technology; SVHN: Street View House Numbers; MLP: Multi-layer perceptron; MCP: Multiplicative Compositional Policies; TCP: Transmission Control Protocol; NLL: Negative Log Likelihood; MCDropout: Monte-Carlo Dropout.

Acknowledgements

None.

Authors contributions

HAD was responsible for conception, design, methods, analysis, drafting, and correspondence.

Funding

None.

Availability of data and materials

Not Applicable.

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

None.

Received: 7 September 2021 Accepted: 15 May 2022
Published online: 31 May 2022

References

1. Kalgaonkar K, Liu C, Gong Y, Yao K. Estimating confidence scores on ASR results using recurrent neural networks. In: 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2015. p. 4999–5003.
2. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 427–436.
3. Chen T, Navrátil J, Iyengar V, Shanmugam K. Confidence scoring using whitebox meta-models with linear classifier probes. In: The 22nd international conference on artificial intelligence and statistics. PMLR; 2019. p. 1467–1475.
4. DeVries T, Taylor GW. Learning confidence for out-of-distribution detection in neural networks. arXiv preprint arXiv:1802.04865. 2018.
5. Erhan D, Szegedy C, Toshev A, Anguelov D. Scalable object detection using deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 2147–2154.
6. Kumar A, Sarawagi S, Jain U. Trainable calibration measures for neural networks from kernel mean embeddings. In: InInternational conference on machine learning. PMLR; 2018. p. 2805–2814.
7. Kastanos A, Ragni A, Gales MJ. Confidence estimation for black-box automatic speech recognition systems using lattice recurrent neural networks. In: ICASSP 2020–2020 IEEE international conference on acoustics, speech and signal processing (ICASSP) . IEEE; 2020. p. 6329–6333.
8. Huang PS, Kumar K, Liu C, Gong Y, Deng L. Predicting speech recognition confidence using deep learning with word identity and score features. In: 2013 IEEE international conference on acoustics, speech and signal processing. IEEE; 2013. p. 7413–7417.
9. Sen PC, Hajra M, Ghosh M. Supervised classification algorithms in machine learning: a survey and review emerging technology in modelling and graphics. Singapore: Springer; 2020. p. 99–111.
10. Thulasidasan S, Chennupati G, Bilmes J, Bhattacharya T, Michalak S. On mixup training: Improved calibration and predictive uncertainty for deep neural networks. arXiv preprint arXiv:1905.11001. 2019.
11. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. 2016. arXiv preprint arXiv:1611.03530. 2017.
12. Otoum S, Kantarci B, Mouftah H. A comparative study of ai-based intrusion detection techniques in critical infrastructures. arXiv preprint arXiv:2008.00088. 2020.
13. Tran NN, Sarker R, Hu J. An approach for host-based intrusion detection system design using convolutional neural network. In: InInternational conference on mobile networks and management. Cham: Springer; 2017. p. 116–126.
14. Demertzis K, Iliadis L, Bougoudis I. Gryphon: a semi-supervised anomaly detection system based on one-class evolving spiking neural network. Neural Comput Appl. 2020;32(9):4303–14.
15. Teyou D, Kamdem G, Ziazet J. Convolutional neural network for intrusion detection system in cyber-physical systems. arXiv preprint arXiv:1905.03168. 2019.
16. Liang S, Li Y, Srikant R. Enhancing the reliability of out-of-distribution image detection in neural networks. arXiv preprint arXiv:1706.02690. 2017.
17. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: International conference on machine learning. PMLR; 2017. p. 1321–1330.
18. Hecker S, Dai D, Van Gool L. Failure prediction for autonomous driving. In: 2018 IEEE intelligent vehicles symposium (IV). IEEE; 2018. p. 1792–1799.
19. Jiang X, Osl M, Kim J, Ohno-Machado L. Calibrating predictive model estimates to support personalized medicine. J Am Med Inform Assoc. 2012;19(2):263–74.
20. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning (still) requires rethinking generalization. Commun ACM. 2021;64(3):107–15.
21. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474. 2016.
22. Kendall A, Cipolla R. Geometric loss functions for camera pose regression with deep learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 5974–5983.
23. Rambach JR, Tewari A, Pagani A, Stricker D. Learning to fuse: A deep learning approach to visual-inertial camera pose estimation. In: 2016 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE; 2016. p. 71–76.
24. Gaonkar B, Beckett J, Attiah M, Ahn C, Edwards M, Wilson B, Laiwalla A, Salehi B, Yoo B, Bui AA, Macyszyn L. Eigenrank by committee: Von-Neumann entropy-based data subset selection and failure prediction for deep learning-based medical image segmentation. Med Image Anal. 2021;67:101834.
25. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6(1):1–48.
26. Ren J, Ren R, Green M, Huang X. Defect detection from X-Ray images using a three-stage deep learning algorithm. In: 2019 IEEE canadian conference of electrical and computer engineering (CCECE). IEEE; 2019. p. 1–4.
27. Tanno R, Worrall DE, Kaden E, Ghosh A, Grussu F, Bizzi A, Sotiropoulos SN, Criminisi A, Alexander DC. Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. NeuroImage. 2021;225:117366.
28. Jassim FA, Altaany FH. Image interpolation using kriging technique for spatial data. arXiv preprint arXiv:1302.1294. 2013.

29. Turner D, Lucieer A, Watson C. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. Remote Sens. 2012;4(5):1392–410.
30. Li L, Li Q, Sun S, Lin HZ, Liu WT, Chen PX. Imaging through scattering layers exceeding memory effect range with spatial-correlation-achieved point-spread-function. Opt Lett. 2018;43(8):1670–3.
31. Juneja M, Sandhu PS. Performance evaluation of edge detection techniques for images in spatial domain. Int J Comput theory Eng. 2009;1(5):614.
32. Chi S, Caldas CH, Kim DY. A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques. Computer-Aided Civ Infrastruct Eng. 2009;24(3):199–211.
33. Hussain AJ, Al-Fayadh A, Radi N. Image compression techniques: A survey in lossless and lossy algorithms. Neurocomputing. 2018;300:44–69.
34. Uthayakumar J, Elhoseny M, Shankar K. Highly reliable and low-complexity image compression scheme using neighborhood correlation sequence algorithm in WSN. IEEE Trans Reliab. 2020;69(4):1398–423.
35. Sujitha B, Parvathy VS, Lydia EL, Rani P, Polkowski Z, Shankar K. Optimal deep learning-based image compression technique for data transmission on industrial Internet of things applications. Trans Emerg Telecommun Technol. 2020;32(7):e3976.
36. Kumar P, Parmar A. Versatile approaches for medical image compression: a review. Procedia Comput Sci. 2020;167:1380–9.
37. Yang J, Bhattacharya K. Combining image compression with digital image correlation. Exp Mech. 2019;59(5):629–42.
38. Rippel O, Bourdev L. Real-time adaptive image compression. In: International conference on machine learning. PMLR; 2017. p. 2922–2930.
39. Kaur A, Gupta S, Sahi L, Padda S. Comprehensive, study of image compression techniques J Crit Reviews. 2020;7(17):2382–8.
40. Johnston N, Eban E, Gordon A, Ballé J. Computationally efficient neural image compression. arXiv preprint arXiv:1912.08771. 2019.
41. LeCun Y. The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/ 1998.
42. Netzer Y, Wang T, Coates A, Bissacco A, Wu B, Ng AY. Reading digits in natural images with unsupervised feature learning, 2011.
43. Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. 2009.
44. Brostow GJ, Fauqueur J, Cipolla R. Semantic object classes in video: A high-definition ground truth database. Pattern Recognit Lett. 2009;30(2):88–97.
45. Jiang H, Kim B, Guan MY, Gupta M. To trust or not to trust a classifier. arXiv preprint arXiv:1805.11783. 2018.
46. 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.
47. Lin P, Zheng C, Yang Y, Zhang F, Yan X. A probability model-based level set method for biomedical image segmentation. J X-Ray Sci Technol. 2005;13(3):117–27.
48. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell. 2012;34(11):2274–82.
49. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. Cham: Springer; 2016. p. 424–432.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.