Computing and Informatics, Vol. 43, 2024, 438–457, doi: 10.31577/cai_2024_2_438

INTERPRETABLE RISK ASSESSMENT METHODS FOR MEDICAL IMAGE PROCESSING VIA DYNAMIC DILATED CONVOLUTION AND A KNOWLEDGE BASE ON LOCATION RELATIONS

Yunan Shi

School of Computer Science and Communication Engineering
Jiangsu University
Zhenjiang, 212013, China
e-mail: 2212108008@stmail.ujs.edu.cn

Junxian Bao

School of Medicine
Jiangsu University
Zhenjiang, 212001, China
&
Zhenjiang Traditional Chinese Medicine Hospital
Zhenjiang, 212003, China
e-mail: 415176750@qq.com

Keyang Cheng*, Weijie Shen

School of Computer Science and Communication Engineering
Jiangsu University
Zhenjiang, 212013, China
e-mail: kycheng@ujs.edu.cn, weijieshen@stmail.ujs.edu.cn

Jingfeng Tang, Yongzhao Zhan

Jiangsu Provincial Key Laboratory of Industrial Network Security Technology
Jiangsu University
Zhenjiang, 212013, China
e-mail: 836221850@qq.com, yzzhan@ujs.edu.cn

* Corresponding author

Abstract. Existing approaches to image risk assessment start with the uncertainty of the model, yet ignore the uncertainty that exists in the data itself. In addition, the decisions made by the models still lack interpretability, even with the ability to assess the credibility of the decisions. This paper proposes a risk assessment model that unites a model, a sample and an external knowledge base, which includes: 1. The uncertainty of the data is constructed by masking the different decision-related parts of the image data with a random mask of probabilities. 2. A dynamically distributed dilated convolution method based on random directional field perturbations is proposed to construct the uncertainty of the model. The method evaluates the impact of different components on the decisions within the local region by locally perturbing the attention region of the dilated convolution. 3. A triadic external knowledge base with relative interpretability is presented to reason and validate the model's decisions. The experiments are implemented on the dataset of CT images of the stomach, which shows that our proposed method outperforms current state-of-the-art methods.

Keywords: High risk areas, quantification of uncertainty, deep learning, dilated convolution, image segmentation, credibility learning

Mathematics Subject Classification 2010: 68T07

1 INTRODUCTION

In recent years, medical image has become a popular area for deep learning research and application. Although "feature learning", represented by deep learning, allows computers to automatically find high-dimensional relevant feature values of targets based on big data, thus achieving fully automated intelligent processing to complete tasks such as target detection, segmentation [1] and prediction in specified application scenarios [2].

Medical diagnosis is fraught with uncertainty. For example, the main features of a patient's condition are the same from doctor to doctor, but different doctors use additional secondary features to aid their diagnosis based on their own experience and accumulated knowledge; imaging doctors have different habits of marking and outlining tissue when reviewing films. These human factors constitute a degree of uncertainty that hinders the integration of data between different doctors and different hospitals. The sharing of data is not always possible. As the samples are not from the same source, training to fit data from different sources results in mutual exclusion, and the model ultimately sacrifices recognition performance on the original source samples in order to enhance generalisation.

Also, the prediction results given by deep learning models are only sometimes reliable. In high-risk areas such as medical imaging and diagnosis, relying exclusively on deep models for decision-making could lead to disastrous consequences [3].

Initially, to equip deep learning models with the ability to judge the plausibility of prediction results, researchers have conducted uncertainty studies around the distribution of data, models, and labels, exploring the impact of various scenarios on model decisions. There are still many gaps in this area of research. Most previous studies have produced different results in the presence of observational uncertainty. Although some researchers have analyzed the factors that generate uncertainty, they have yet to use the uncertainty rationally to improve the model's performance.

Furthermore, since the depth model is essentially an end-to-end black box model [4], we think that there are limitations in assessing the uncertainty of a model singularly, as reflected in the fact that the source of confidence is only an isolated task, with no external means to validate and support it.

To address the above limitations, this paper proposes a model for jointly constructing quantitative risk assessments of diagnostic decisions with multiple levels of uncertainty. The model combines data uncertainty and sensory uncertainty to synthetically assess decision outcomes, and ultimately queries and tests the credibility of decisions through a knowledge base of positional relationships.

In summary, our main contributions are:

- A probability-based random mask noise is applied to mask non-target regions from the part of the data negatively correlated with the decision. The final decision is optimized using a combination of uncertain optimization results.
- A randomly perturbed dynamically dilated convolution is proposed to allow the model to make diverse decisions based on the characteristics of different distributions.
- The concept of knowledge mapping is transferred to images. The gastric cavity, the tumour, and the related phase information are employed to construct a comprehensible external knowledge base, thus making the decision-making process transparent and interpretable.

2 RELATED WORK

In recent years, researchers have measured decision uncertainty in terms of models, data, and labels. In 2015, Gal and Ghahramani proposed a Dropout method based on Bayesian probability to capture the uncertainty of the model [5]. The method measures model uncertainty by creating randomness in the model's parameters so that we can capture the decisions made by the model under different circumstances. However, the method is limited by the shift in the dataset. To address the impact of shifting, Fort et al. proposed the Deep Ensembles model [6] in 2019. This model captures the uncertainty of the model by adjusting the degree of training each time so that the model obtains the optimal solution for different locals. Considering the visual ambiguity, Kohl et al. introduced conditional variational self-encoders in U-Net to form a segmentation model that generates as many hypotheses as possible [7]. In 2022, Guo et al. built label correlation networks [8] from label distribution

to model the label relation uncertainty. The majority of the methods above use random probabilities to generate uncertainty. Although uncertainties are analyzed to some extent, they are not exploited to enhance current models. This paper explores this topic as a result.

2.1 Model Validation

To evaluate the reliability of a model and verify its performance. Initially, Seymour Geisser proposed the cross-validation method [9], which divided the dataset into a training set and a validation set for training and validation, respectively. However, this method was too random, and the results needed more convincing. The emergence of K-fold Cross Validation [10] has solved this problem to some extent. The method uses different groupings to train K models and combines the K models to obtain more convincing results. With the rapid development of statistics, Maurice Quenouille proposes a resampling method for generic hypothesis testing and confidence interval calculation [11] to reduce the bias of the estimates.

As an end2end posterior distribution prediction black box model, once the deep model loses the support of actual labels, it will be difficult to assess the reliability of the model's decisions and lacks persuasive power. This is why deep models have been studied on a large scale in high-risk areas but need to be put into practical use. In order to remedy the problems raised above, this paper proposes a cross-domain backtesting approach with decision-making in the high-risk domain of medical diagnosis.

2.2 Interpretable Methods of Semantics

Explanatory methods can generally be divided into Post-hoc reasoning and Pre-hoc reasoning. Post-hoc explanation represents a unique approach to extract information from a learned model. Although the working principle of the model cannot be accurately elucidated, for a given trained distributed inference model, certain explanation of the model's working mechanism, decision-making behavior, and evidence basis can be made by using explanation methods or constructing explanation models, such as CAM [12], Grad-CAM [13] and Score-CAM [14], which are all applied to understand and reason the behavior of the network through visualization. Pre-hoc explanation models refer to models that are inherently interpretable or integrate interpretable modules into their architecture. For a trained learning model, the decision-making process or decision basis of the model can be understood without additional information, such as knowledge graph, which can explain the model's decision through queries of existing libraries.

Most of the existing interpretable models and methods for interpreting them have been studied in classification tasks and are difficult to transfer to tasks such as segmentation, target detection, etc. Our work is inspired by knowledge graphs and utilizes the gastric cavity, the tumour, and the related phase information to build

triadic knowledge graphs with relative interpretability, exploiting the knowledge of the library to query and validate the decisions of the model.

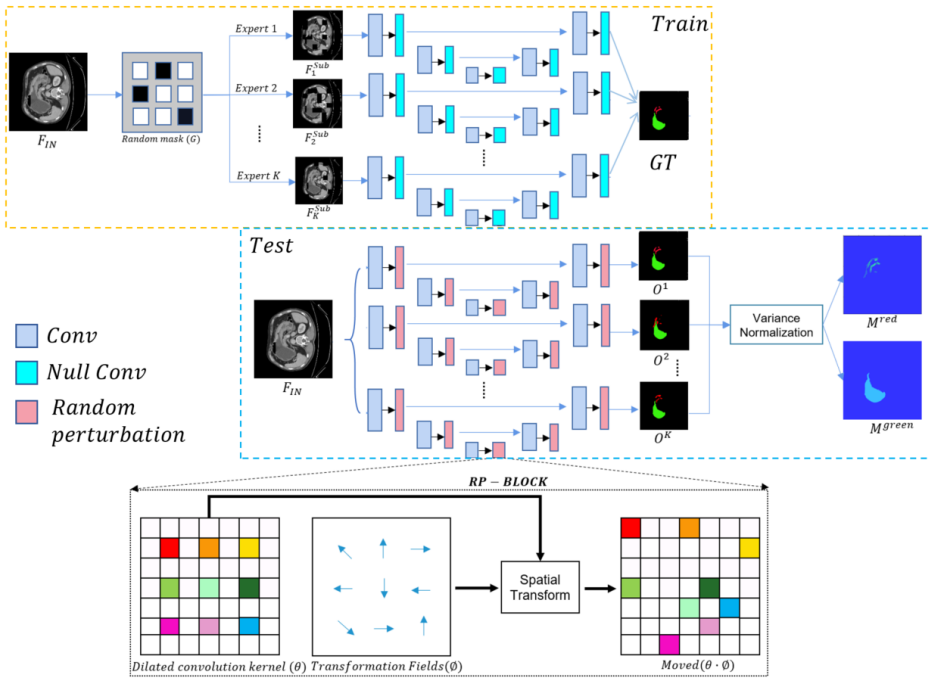

Figure 1. The overall framework structure of the Dynamic Unet

3 PROPOSED METHOD

The Dynamic Unet proposed in this paper is shown in Figure 1, which mines decision uncertainty from both data and sensory aspects, respectively. Firstly, Dynamic Unet introduces random masks in the training phase, which obtains K mutually exclusive subsets F_i^{Sub} of the input F_{IN}. The method makes full use of data from different components of the same image to train K different expert models θ_i.

Secondly, the features that different people focus on when identifying the same thing vary from person to person. Therefore, in the testing phase, a dynamically distributed null convolution kernel was added to the model, which obtains K different decision outcomes O_i by perturbing the focal positions of the dilated convolution kernel through random orientation fields. The aim of the method is to simulate human sensory uncertainty.

Furthermore, our experiments model the phase (relationship) association of the gastric lumen (subject), tumour (object), and pathology to train a positional rela-

tionship external knowledge base, as shown in Figure 2. The credibility of model decisions can be queried and tested through this knowledge base.

3.1 Data Uncertainty

Most of the previous uncertainty assessment methods based on the Dropout mechanism were carried out in the testing phase, and although the method constructs the parameter uncertainty of the model by partially inactivating neurons, it sacrifices part of the model's performance. In addition, the Dropout method can only be applied to the fully-connected layer, and cannot be applied precisely to the pixel level.

In deep learning interpretability, the correlation between output and input has previously been visualized utilizing gradient imputation, yielding that different local features of the input image play either a positive or negative correlation role in model fitting. In order to maximise the retention of model performance when building uncertainties, we shift the timing of building uncertainties from the testing phase to the training phase. In this paper, we start from this property and perform random and differential deactivation of noisy data in non-tumour regions in gastric cancer tumours' CT image segmentation task. For a single sample F_{IN}, multiple subsets of samples are obtained by a probability-based mask G. The formula is as follows:

$$G = \{G_i \mid i = 1, 2, 3, \ldots, K\}, \tag{1}$$

$$G_i^{h,w} = \begin{cases} 0, & p_i^{h,w}, \\ 1, & 1 - p_i^{h,w}, \end{cases} \tag{2}$$

$$F_i^{Sub} = G_i * F_{IN}, \tag{3}$$

$$p_{i+1}^{h,w} = \begin{cases} p_i^{h,w} * p, & G_i^{h,w} = 0, \\ p_i^{h,w}, & G_i^{h,w} = 1, \end{cases} \tag{4}$$

where K denotes the number of random samples, $G_i^{h,w}$ represents the local window with index (h, w) and $p_{i+1}^{h,w}$ denotes the probability of each pixel point in the image being masked. The probability decreases exponentially with the number of times the mask covers it. Under this approach, the final n subsets of input samples F_{IN} that are different from each other, i.e. $F_i^{Sub} \in F_{IN}$.

This approach weakens the different attribution components, allowing the model to output different results and combining the different results to assess the risk of the model decisions. As shown in Figure 2, after the noise mask overlays the features at different locations, the determined feature map F_{IN} is separated into several subsets F_i^{Sub} of feature maps. Due to the nature of the subsets being different from each other, each subset can be trained with a personalised model $\{\theta_i \mid F_i^{Sub}, i = 1, 2, \ldots, 8\}$.

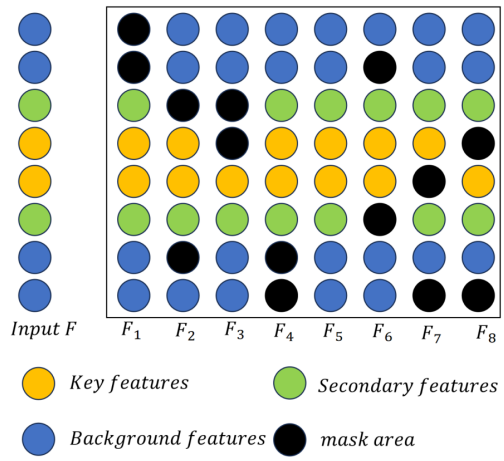

Figure 2. Schematic diagram for constructing attentional uncertainty

3.2 Sensory Uncertainty

Previous Doupout methods could not embed the convolutional layer, and once the convolutional layer was added to the Dropout, it would cause most of the features to be lost. In order to overcome the shortcomings of Dropout, the perturbation-based null convolution kernel proposed in this paper guarantees parameter integrity and constructs uncertainty by locally perturbing the position of the parameters.

The features that people focus on to identify the same thing are different, but rather vary from person to person, which creates sensory uncertainty. Furthermore, instead of focusing on each pixel point, the recognition of an image usually involves applying a few key features as an alternative to the overall features. Therefore, we propose a dilated convolution method with dynamic feature point properties that can be embedded in any deep network. The method simulates sensory uncertainty by randomly shifting the position of key features so that the model can perceive different features to make different decisions, while retaining the benefits of dilated convolution.

To prevent the perturbed dilated convolution kernel from sensing too many edge feature points and to reduce the probability of feature point overlap, we choose to apply a Gaussian distribution with high central probability to control point shifting. The human visual field critical sight is modelled by treating a convolution kernel of size d^2 as d^2 discrete attention critical points, expanding the convolution kernel to form a void convolution with the parameters $rate = r$ and $padding = 0$. With the discrete vital points as the centre and r as the radius to extract a single visual field point window $\{C_i \mid i = 1, 2, 3, \ldots, K\}$. We are constructing binary Gaussian functions using the transverse and longitudinal aspects of the convolution kernel.

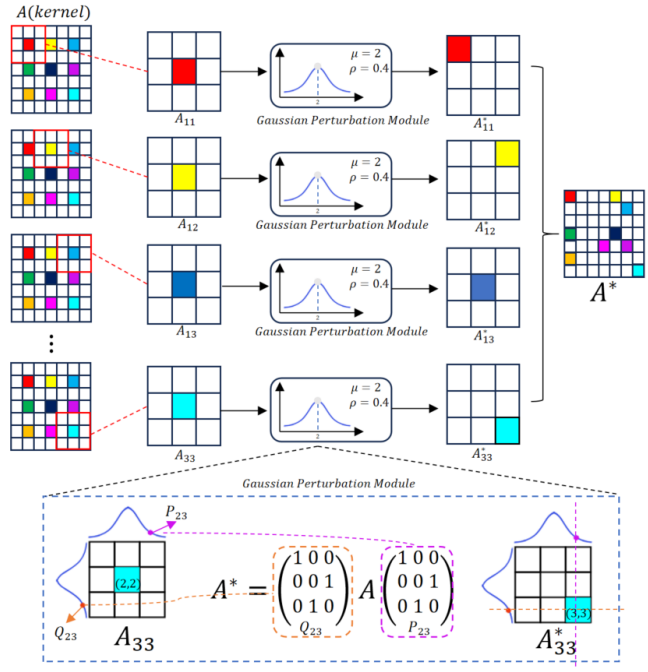

Figure 3. Structure of the Random Perturbation Block

The formula for the binary Gaussian distribution function is as follows:

$$P(X) = \frac{1}{|2\pi\sigma|} \exp\left(-\frac{1}{2}(X-\mu)^T \sigma^{-1}(X-\mu)\right), \tag{5}$$

where $X \sim (x, y)$, x and y denote the horizontal and vertical position coordinates of the window center as the origin, $[\mu, \sigma] = \begin{bmatrix} 0 & r \\ 0 & r \end{bmatrix}$, and the shift of the control key point line of sight is randomly sampled in the probability distribution range $P(X) \sim N(\mu, \sigma)$.

The perturbed null convolution kernel A^* is obtained by shifting the point features through the row transformation matrix and the column transformation matrix after random sampling of $X^* \sim (x^*, y^*)$.

$$A^* = Q \times A \times P, \tag{6}$$

where Q represents the row transformation matrix and P represents the column transformation matrix

After building uncertainty in the data and model layers, a deterministic model is trained into K models with different parameters and different concerns. In the

testing phase, a single input can output different i decisions.

$$O_i = U_i\left(F_i^{Sub}, A^*\right), \tag{7}$$

$$O_S = \frac{\sum_{i=1}^{K}\left(O_i - \left(\frac{\sum_{i=1}^{K}(O_i)}{K}\right)\right)^2}{K}, \tag{8}$$

where $U_i\left(F_i^{Sub}, A^*\right)$ represents the sub-model numbered i adding the perturbation A^*, O_i is the lesion mask corresponding to the output of the model, and O_S represent the mean and variance maps respectively, where the greater the variance of the corresponding location of the pixel point, the greater the risk.

After completing the variance calculation for multiple expert model decisions, the risk assessment task is completed by normalization. The tumour M^{red} and gastric cavity M^{green} are presented by utilizing a heat map.

$$X^* = \frac{X - min}{max - min}, \tag{9}$$

where min and max represent the smallest variance and the largest variance in the matrix, respectively.

3.3 Knowledge Graph Triplet Construction

In the third step, our work proposes to train a relatively interpretable triadic knowledge graph by establishing associations between gastric lumen, tumour and tumour phase information. Ultimately, the model is able to perform a query test on the decision by querying the knowledge graph, as shown in Figure 4.

During the testing phase, additional location information of the obtained gastric cavity segmentation mask and tumour segmentation mask is encoded for extracting features f^{green}, f^{red}. The Ternary knowledge graphs are then utilized to analyze the correlation between the two in order to output the phase category label T_i. The fc layer then increases the dimensionality of the labels, and the added labeled features f_{T_i} and gastric lumen mask features f^{green} containing location information are fed into the knowledge graph, and the model outputs the features f^{pred} after analysing the correlation. Finally, the approximate position of the tumour in the map is output by upsampling.

$$T_i = LSTM^1(f^{green}, f^{red}), \tag{10}$$

$$f_{T_i} = fc(T_i), \tag{11}$$

$$f^{pred} = LSTM^2(f^{green}, f_{T_i}), \tag{12}$$

$$M^{pred} = upsample(f^{pred}), \tag{13}$$

where fc represents a fully connected layer, *upsample* represents the upsampling operation, T_i denotes the location relationship between the gastric lumen and the

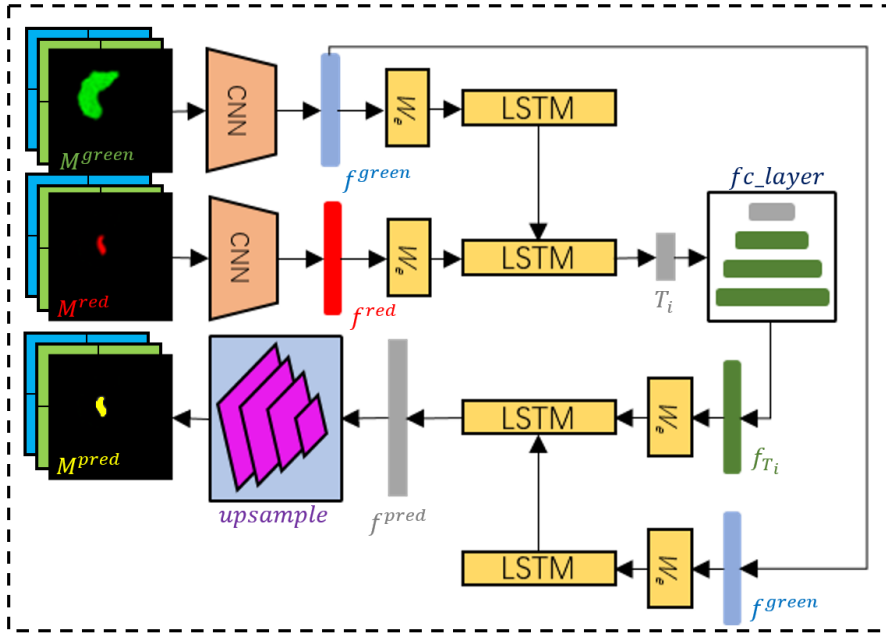

Figure 4. Methodology for constructing a knowledge base of location relations

tumour, T_i = {Lesser curvature of gastric body, Cardia, Cardia and small curved side of gastric body, Gastric antrum, Gastric antrum and small curved side of gastric body}.

Ultimately, the lesion threshold mask map O_f obtained by querying the knowledge graph is compared with the lesion mask map M_i^{pred} from the segmentation task to assess the risk indicators.

$$Risk = 1 - Dice\left(O_f, M_i^{pred}\right). \tag{14}$$

where *Dice* is a metric for semantic segmentation. The equation means that the more the predicted lesion mask is outside the threshold range, the greater the risk.

3.4 Train Loss

The model layer mainly involves segmenting between the tumour and the gastric lumen. In the segmentation task, we introduce a cross-entropy loss function, which is as follows.

$$\ell_{CE}^{1} = -\frac{1}{H \times W} \sum_{i=0}^{H \times W} y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i), \tag{15}$$

where H and W indicate the length and width of the image, y_i represents the predicted label of the current pixel, and $\hat{y}_i$ is defined as the probability that the predicted outcome is the label.

We also apply the cross-entropy function to the prediction of phase labels. The loss function is defined as follows:

$$\ell^2_{CE(g,g^*)} = -\sum_{i=0}^{C-1} g_i \log \hat{g}_i, \tag{16}$$

where C denotes the number of categories, g_i represents the predicted label of the current pixel, and $\hat{g}_i$ is defined as the probability corresponding to the predicted label.

The MSE loss function is defined as follows:

$$\ell_{MSE(t,t^*)} = \frac{\sum_{i=0}^{n}(t_i - t_i^*)}{H \times W}, \tag{17}$$

where t_i represents the predicted value of the pixel point, and t_i^* is defined as the true tumour mask map.

4 EXPERIMENTS

4.1 Parameters and Running Time

The proposed model was built by calling PyTorch in the Python 3.6 environment. The experimental platform is i7-6950 + 4 × RTX1080Ti, the memory is 128 GB, and the operating system is Ubuntu 19.01. A positive-terrestrial distribution is used to initialize the parameters of the feature layer of the model. It uses the Adam optimizer with a learning rate between 0.001 and 0.0001. The pre-training phase of the proposed GDCNet master network is carried out using medical stomach cancer CT image maps, lasting 11 hours and requiring approximately 7 minutes for each epoch.

4.2 Dataset

This section evaluates the proposed method on two medical image segmentation datasets.

CT-GC contains 500 CT serial slice images of gastric lesions from the First People's Hospital of Zhenjiang City, Jiangsu Province and the corresponding medical report of the case. All images in the dataset are grayscale images with a resolution of 512 × 512, which are stored in DICOM format. Several imaging specialists at the same hospital manually annotated the lesion labels for this dataset. We randomly select 80 % of the dataset (i.e. 400 cases) as our training set and the rest as the test set.

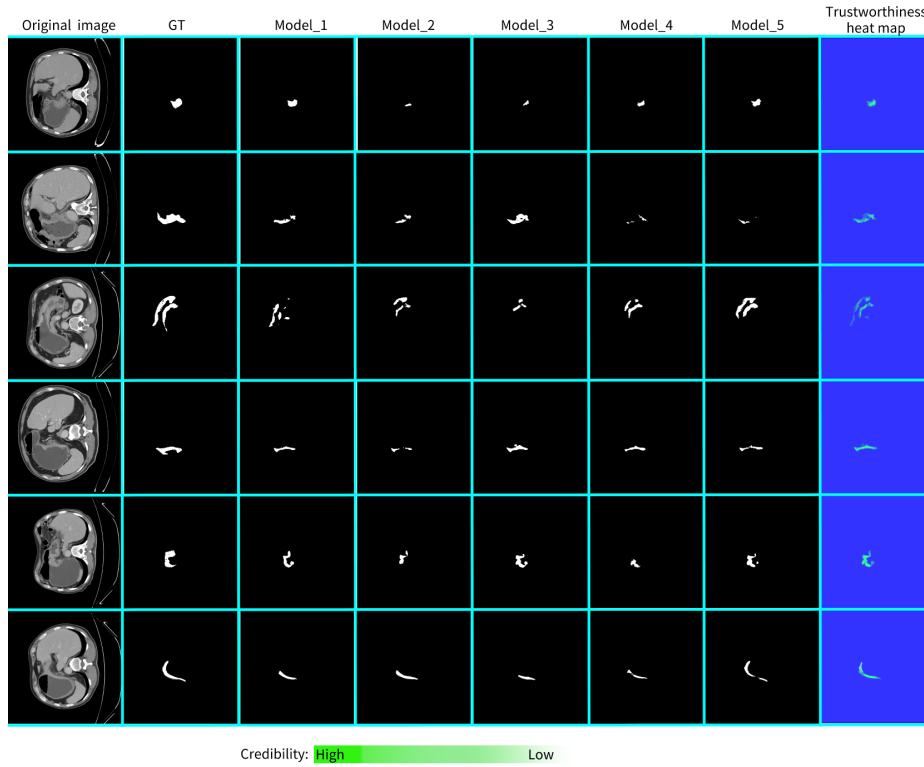

Figure 5. Graph of experimental results for the Dynamic Unet

STARE dataset is a project initiated by Michael Goldbaum in 1975 and first published in 2000 by Hoover et al. It is a colour fundus image database for retinal vessel segmentation, including 20 fundus images, 10 with and 10 without lesions, with a resolution of 605×700, and each image corresponding to two manual segmentation results by experts. It is one of the most commonly used standard fundus image libraries.

4.3 Evaluation Metrics

In evaluating the performance of the image segmentation stage models, we assessed the accuracy of image segmentation using $DICE$, Auc, and $Recall$ commonly used in medical segmentation tasks.

The following equation gives the definition of $Dice$:

$$scale = 0.25Dice = \frac{2|P \cap G|}{|P| + |G|}, \tag{18}$$

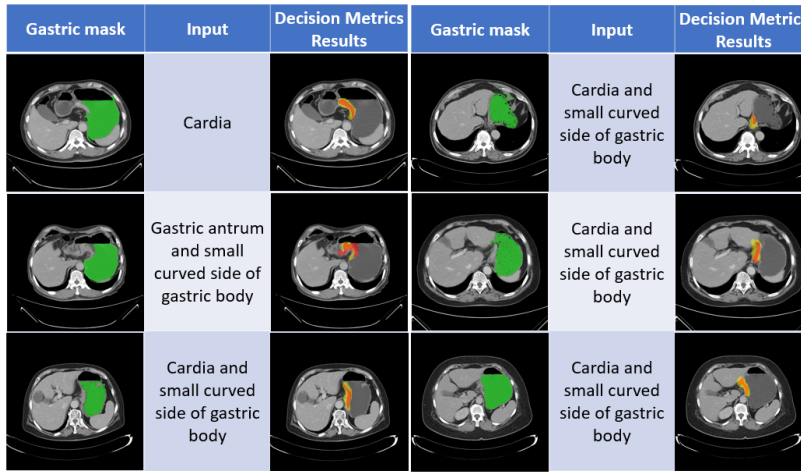

Figure 6. Diagram of the effect of query checking of the location relationship knowledge base

where $|\cdot|$ denotes the cardinality of a set, P stands for the network prediction and G is the ground truth.

Precision is denoted by:

$$Auc = \frac{TP + TN}{TP + FP + TN + FN}, \tag{19}$$

Dataset	Method	Dice	AUC	Recall
CT-GC	MH [15]	0.721	0.751	0.741
	UNet Ensemble [16]	0.703	0.718	0.698
	I2I [17]	0.587	0.545	0.603
	ProbU-Net [7]	0.612	0.578	0.581
	PHISeg [3]	0.604	0.512	0.594
	DroUNet [18]	0.691	0.645	0.713
	LDLV [8]	0.679	0.734	0.705
	DUnet (ours)	0.795	0.831	0.807
STARE	MH [15]	0.631	0.611	0.586
	UNet Ensemble [16]	0.688	0.731	0.751
	I2I [17]	0.641	0.637	0.654
	ProbUNet [7]	0.676	0.709	0.722
	PHISeg [3]	0.733	0.741	0.759
	DroUNet [18]	0.656	0.698	0.704
	LDLV [8]	0.749	0.757	0.708
	DUnet (ours)	0.751	0.766	0.731

Table 1. Comparison of experimental results

where TP, FP stand for true positive and false positive, respectively, and TN, FN stand for true negative and false negative, respectively.

Recall is calculated by:

$$Recall = \frac{TP}{TP + FN}, \tag{20}$$

where TP, FN stand for true positive and false negative, respectively.

4.4 Results and Comparisons

As shown in Figure 5, we have visualised the output of Dynamic Unet, from which we can observe that the five expert models produce different results based on different features of the images. The results of these five decisions were variance and normalised and finally presented in the form of a heat map. Rather than simply overlaying the results of the five decisions, the heat map additionally provides the confidence level of each pixel point after the decision has been classified. The darker the colour, the higher the confidence level. It is worth noting that the variance here can be replaced by the information entropy, both of which measure the uniformity of the five expert models' decisions.

In addition, we compare our approach with four classical models (Dropout U-Net [18], U-Net Ensemble [16], M-Heads [15], and Image2Image VAE [17]) and three state-of-the-art approaches (PHISeg [3], Probabilistic U-Net [7] and LDLV GRID [8]). We have replicated the model with the ideas provided by the authors. For all comparisons, we have kept the default settings suggested by the authors as much as possible.

Finally, our experiments visualised the queries from the location-relational knowledge base. As shown in Figure 6, the red mask is from the segmentation model, while the yellow mask indicates the overall extent of lesion occurrence at the corresponding location, meaning that the decision is more risky when the red part exceeds the threshold.

The complete model achieves the best results on both the *Dice* and *AUC* metrics, as well as a good performance on the *Recall* metric, thus validating the effectiveness of the proposed model.

4.5 Ablation Study

4.5.1 Dynamic Cavity Convolution

To verify the validity of our proposed model, we designed a baseline and evaluated five variants of the baseline. The baseline is the Dynamic Unet proposed in the paper. The five variants evaluate the model's performance in terms of the number of K-sample and the size of the dilated convolutional rate. After extensive experiments, as shown in Figure 7, the best performance is achieved when Sample-$K = 6$, $rate = 5$.

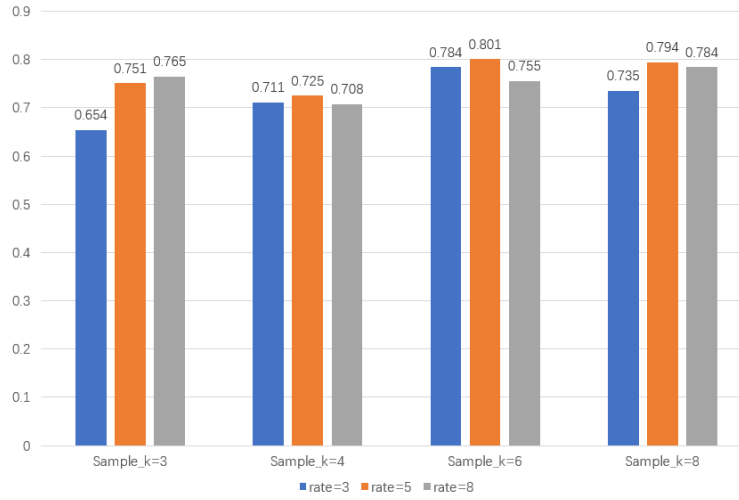

Figure 7. Ablation experiment histogram

Statistical analysis: It is concluded that the performance of the model shows a steady trend of improvement as the value of K is taken to increase. Theoretically, the performance of the model converges to a stable value when the value of K tends to positive infinity, i.e. when the set of features reaches an upper limit. In addition, the size of the dynamic hole convolution needs to be controlled in the process of increasing the number of K branches. When the size of the dynamic hole convolution exceeds a suitable threshold, it increases the potential of losing key features of the input data, affecting the final performance of the model.

4.5.2 Random Perturbation

In order to illustrate why the Gaussian function is used as a means of random perturbation and to analyse its superiority, in this section we perform ablation experiments with random Gaussian and random uniform distributions.

As shown in Figure 8, the perturbation method based on the binary Gaussian distribution significantly outperforms the uniform distribution in the *Dice* metric as the rate grows, and the number of random samples increases.

Statistical analysis: Observing the figure, it can be noticed that at $rate = 3$ i.e. when the space to move the keypoints is relatively small, the performance difference between the two is not significant. Once the moveable space increases, the perturbation based on the mean function makes the convolution kernel more cheap with invalid features at the edges, which greatly affects the performance of the model. Gaussian function based perturbation will focus more on the main features related to the centre.

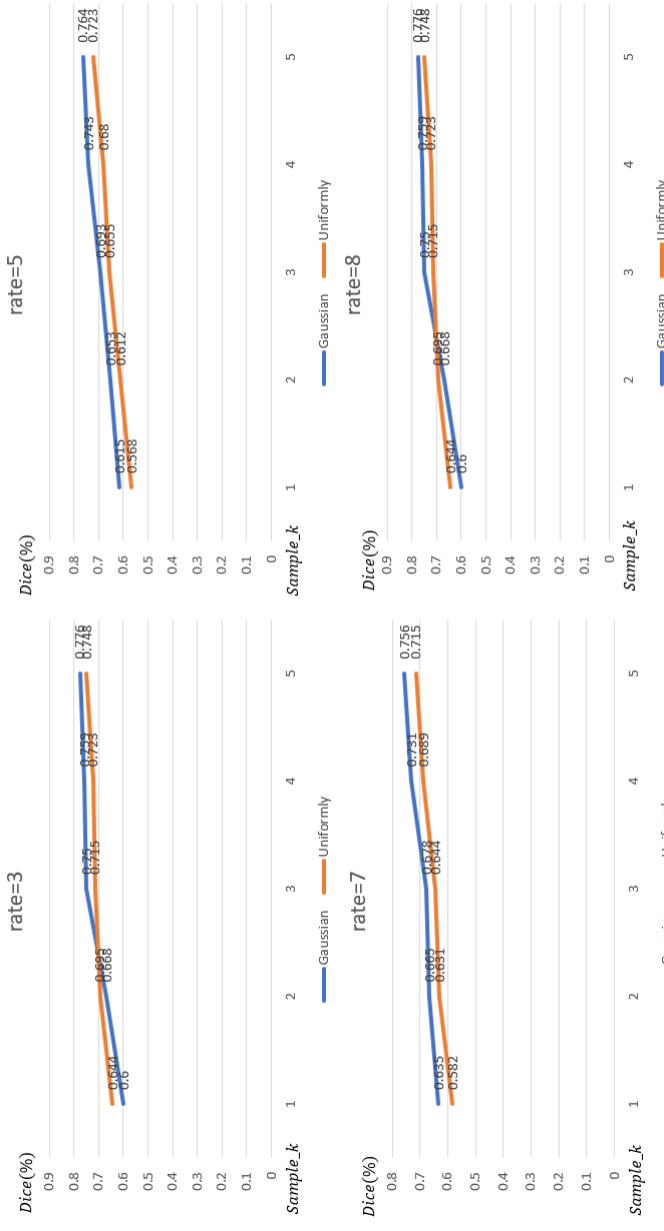

Figure 8. Folded graph of the ablation experiment

5 CONCLUSION

This paper proposes a new Dynamic Unet for assessing the degree of risk in model decisions. First, the uncertainty of the training data is created by randomly masking the noise to obtain a set of mutually dissimilar, expert models. Secondly, simulating sensory uncertainty by introducing a dynamically expanding convolutional kernel allows the model to perceive different features to make personalized decisions. Thirdly, a positional relationship base is constructed to assess the credibility of the decisions, making the model's decisions relatively interpretable using the inference mechanism of the knowledge base. Ultimately, the validity of the proposed method was demonstrated by a quantitative and qualitative study of a medical CT dataset of gastric cancer.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 62372215) and Jiangsu Province Traditional Chinese Medicine Technology Development Plan Project (No. MS02140) and Special Fund Project of Jiangsu Science and Technology Plan (No. BE2022781).

REFERENCES

[1] Zhang, J.—Xie, Y.—Xia, Y.—Shen, C.: DoDNet: Learning to Segment Multi-Organ and Tumors from Multiple Partially Labeled Datasets. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 1195–1204, doi: 10.1109/CVPR46437.2021.00125.

[2] Chiu, Y. C.—Zheng, S.—Wang, L. J.—Iskra, B. S.—Rao, M. K.—Houghton, P. J.—Huang, Y.—Chen, Y.: Predicting and Characterizing a Cancer Dependency Map of Tumors with Deep Learning. Science Advances, Vol. 7, 2021, No. 34, Art. No. eabh1275, doi: 10.1126/sciadv.abh1275.

[3] Baumgartner, C. F.—Tezcan, K. C.—Chaitanya, K.—Hötker, A. M.—Muehlematter, U. J.—Schawkat, K.—Becker, A. S.—Donati, O.—Konukoglu, E.: PHiSeg: Capturing Uncertainty in Medical Image Segmentation. In: Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P. T., Khan, A. (Eds.): Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer, Cham, Lecture Notes in Computer Science, Vol. 11765, 2019, pp. 119–127, doi: 10.1007/978-3-030-32245-8_14.

[4] Zhang, Q. S.—Zhu, S. C.: Visual Interpretability for Deep Learning: A Survey. Frontiers of Information Technology & Electronic Engineering, Vol. 19, 2018, No. 1, pp. 27–39, doi: 10.1631/FITEE.1700808.

[5] Gal, Y.—Ghahramani, Z.: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Proceedings of the 33rd International Conference on Machine Learning, Proceedings of Machine Learning Research (PMLR), Vol. 48, 2016, pp. 1050–1059, http://proceedings.mlr.press/v48/gal16.pdf.

[6] Fort, S.—Hu, H.—Lakshminarayanan, B.: Deep Ensembles: A Loss Landscape Perspective. CoRR, 2019, doi: 10.48550/arXiv.1912.02757.

[7] Kohl, S.—Romera-Paredes, B.—Meyer, C.—De Fauw, J.—Ledsam, J. R.—Maier-Hein, K.—Ali Eslami, S. M.—Jimenez Rezende, D.—Ronneberger, O.: A Probabilistic U-Net for Segmentation of Ambiguous Images. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 31 (NeurIPS 2018). Curran Associates, Inc., 2018, pp. 6965–6975, https://proceedings.neurips.cc/paper_files/paper/2018/file/473447ac58e1cd7e96172575f48dca3b-Paper.pdf.

[8] Guo, Q.—Zheng, Z.—Jia, X.—Xu, L.: Label Distribution Learning via Label Correlation Grid. CoRR, 2022, doi: 10.48550/arXiv.2210.08184.

[9] Refaeilzadeh, P.—Tang, L.—Liu, H.: Cross-Validation. Encyclopedia of Database Systems, Springer US, Vol. 5, 2009, pp. 532–538.

[10] Fushiki, T.: Estimation of Prediction Error by Using K-Fold Cross-Validation. Statistics and Computing, Vol. 21, 2011, No. 2, pp. 137–146, doi: 10.1007/s11222-009-9153-8.

[11] Yu, C. H.: Resampling Methods: Concepts, Applications, and Justification. Practical Assessment, Research, and Evaluation, Vol. 8, 2019, No. 1, Art. No. 19, doi: 10.7275/9cms-my97.

[12] Zhou, B.—Khosla, A.—Lapedriza, A.—Oliva, A.—Torralba, A.: Learning Deep Features for Discriminative Localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2921–2929, doi: 10.1109/CVPR.2016.319.

[13] Selvaraju, R. R.—Cogswell, M.—Das, A.—Vedantam, R.—Parikh, D.—Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[14] Wang, H.—Wang, Z.—Du, M.—Yang, F.—Zhang, Z.—Ding, S.—Mardziel, P.—Hu, X.: Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 24–25, doi: 10.1109/CVPRW50498.2020.00020.

[15] Lee, S.—Purushwalkam, S.—Cogswell, M.—Crandall, D.—Batra, D.: Why M Heads Are Better Than One: Training a Diverse Ensemble of Deep Networks. CoRR, 2015, doi: 10.48550/arXiv.1511.06314.

[16] Lakshminarayanan, B.—Pritzel, A.—Blundell, C.: Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. In: Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates, Inc., 2017, pp. 6402–6413, https://proceedings.neurips.cc/paper_files/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf.

[17] Isola, P.—Zhu, J. Y.—Zhou, T.—Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Com-

puter Vision and Pattern Recognition (CVPR), 2017, pp. 5967–5976, doi: 10.1109/CVPR.2017.632.

[18] Guo, C.—Szemenyei, M.—Pei, Y.—Yi, Y.—Zhou, W.: SD-Unet: A Structured Dropout U-Net for Retinal Vessel Segmentation. 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), 2019, pp. 439–444, doi: 10.1109/BIBE.2019.00085.

Yunan Shi received his Bachelor's degree in electrical engineering from the Huaiyin Normal University, Huaiyin, China, in 2021. He is currently a postgraduate student in the Department of Computer Science at the Jiangsu University. His current research interests are in computer vision, medical image processing, and NLP.

Junxian Bao received her Master's degree in clinical medicine from the Jiangsu University and is now working at the Zhenjiang Chinese Hospital.

Keyang Cheng received his Ph.D. from the Nanjing University of Aeronautics and Astronautics in 2010 and his postdoctoral studies at the University of Warwick in the UK in 2016. He is now Executive Vice President of the Institute of Cyberspace Security, Professor and Doctoral Fellow at the Jiangsu University.

Weijie Shen received his Bachelor's degree in software engineering from the Jinling Institute of Technology, Nanjing, China, in 2021. He is currently a Master student at the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China. His research interests include computer vision, medical image, and their applications.

Jingfeng Tang received his B.Sc. degree in Internet of Things engineering from the Huaiyin Normal University, Huaiyin, China, in 2020. He is currently a Master candidate at the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China. His research interests include artificial intelligence, computer vision.

Yongzhao Zhan received his Ph.D. from the Department of Computer Science and Technology at the Nanjing University in 2000, and is currently Professor and doctoral supervisor in the School of Computer Science and Technology at the Jiangsu University.