Uncertainty-aware Multi-modal Learning via Cross-modal Random Network Prediction

H Wang , J Zhang , Y Chen , C Ma , J Avery, L Hull and G Carneiro

Abstract

We propose a new Uncertainty-aware Multi-modal Learner that estimates uncertainty by measuring feature density via Cross-modal Random Network Prediction (CRNP). CRNP is designed to require little adaptation to translate between different prediction tasks, while having a stable training process. From a technical point of view, CRNP is the first approach to explore random network prediction to estimate uncertainty and to combine multi-modal data. Experiments on two 3D multi-modal medical image segmentation tasks and three 2D multi-modal computer vision classification tasks show the effectiveness, adaptability and robustness of CRNP.

Results 1

Table 1: The performance comparison of CRNP and different challenge models on both CT and MR segmentation of MMWHS dataset. The best results for each column are in bold. ↑ sign indicates the higher value the better.

	CT		MR	
Models	Dice ↑	Jaccard ↑	Dice ↑	Jaccard ↑
GUT	0.9080	0.8320	0.8630	0.7620
KTH	0.8940	0.8100	0.8550	0.7530
CUHK1	0.8900	0.8050	0.7830	0.6530
CUHK2	0.8860	0.7980	0.8100	0.6870
UCF	0.8790	0.7920	0.8180	0.7010
SIAT	0.8490	0.7420	0.6740	0.5320
UT	0.8380	0.7420	0.8170	0.6950
UB1	0.8870	0.7980	0.8690	0.7730
UB2	-	-	0.8740	0.7780
UOE	0.8060	0.6970	0.8320	0.7200
Ours	**0.9193**	**0.8486**	**0.8758**	**0.7814**

We compare the proposed CRNP model with the state-of-the-art models reported by the official challenge report [1]. The results are shown in the table. On whole heart segmentation, CRNP has a particularly accurate Dice score and Jaccard index for CT and MR. Compared to the second-best models, our CRNP model increases the Dice score from 0.9080 to 0.9193 and from 0.8740 to 0.8758 on CT and MR, respectively. Similar results are shown forJaccard index.

[1] Xiahai Zhuang et al. Evaluation of algorithms for multi-modality whole heart segmentation. *Medical image analysis*, 2019.

Contributions

- We propose a new uncertainty-aware multi-modal learning model through a feature distribution learner named as Cross-modal Random Network Prediction (CRNP). CRNP is designed to be easily adapted to disparate tasks and to be robust to numerical instabilities during optimization.
- We introduce a novel uncertainty estimation based on fitting the output of an RNP, which from a technical viewpoint, represents a departure from more common uncertainty estimation methods based on Bayesian learning or abstention mechanisms.

Results 2

Table 2: The performance of different models on BraTS2020 Online validation set. The best results for each column are in bold. ∗ indicates models with ensemble. ↑ sign indicates the higher value the better; while ↓ means the lower value the better.

	Dice ↑			Hausdorff95 ↓		
Models	ET	WT	TC	ET	WT	TC
3D UNet [6]	0.6876	0.8411	0.7906	50.9830	13.3660	13.6070
Basic VNet [25]	0.6179	0.8463	0.7526	47.7020	20.4070	12.1750
Deeper VNet [25]	0.6897	0.8611	0.7790	43.5180	14.4990	16.1530
Residual 3D UNet	0.7163	0.8246	0.7647	37.4220	12.3370	13.1050
ProbUNet [19]	0.7392	0.8782	0.7955	36.2458	6.9518	7.7183
SSN [26]	0.6795	0.8420	0.7866	43.6574	14.6945	19.5171
Modal-Pairing* [40]	0.7850	0.9070	0.8370	35.0100	4.7100	5.7000
TransBTS [38]	0.7873	0.9009	0.8173	**17.9470**	4.9640	9.7690
CRNP (Ours)	0.7887	0.9086	0.8372	26.5972	**4.0490**	6.0040
CRNP* (Ours)	**0.7902**	**0.9109**	**0.8550**	26.4682	4.1096	**5.3337**

Developing automated segmentation models to delineate intrinsically heterogeneous brain tumors is the main goal of BraTS2020 Challenge. We compare the proposed CRNP model with many other strong methods, including 3D UNet, Basic VNet, Deeper VNet, Residual 3D UNet, Modal-Pairing, TransBTS, as well as uncertainty-aware models ProbUNet and SSN that models aleatoric uncertainty by considering spatially coherence. We evaluate the Dice and Hausdorff95 indexes of all models on four organs: enhancing tumor (ET); tumor core (TC) that consists of ET, necrotic and nonenhancing tumor core; and whole tumor (WT) that contains TC and the peritumoral edema.

Methodology

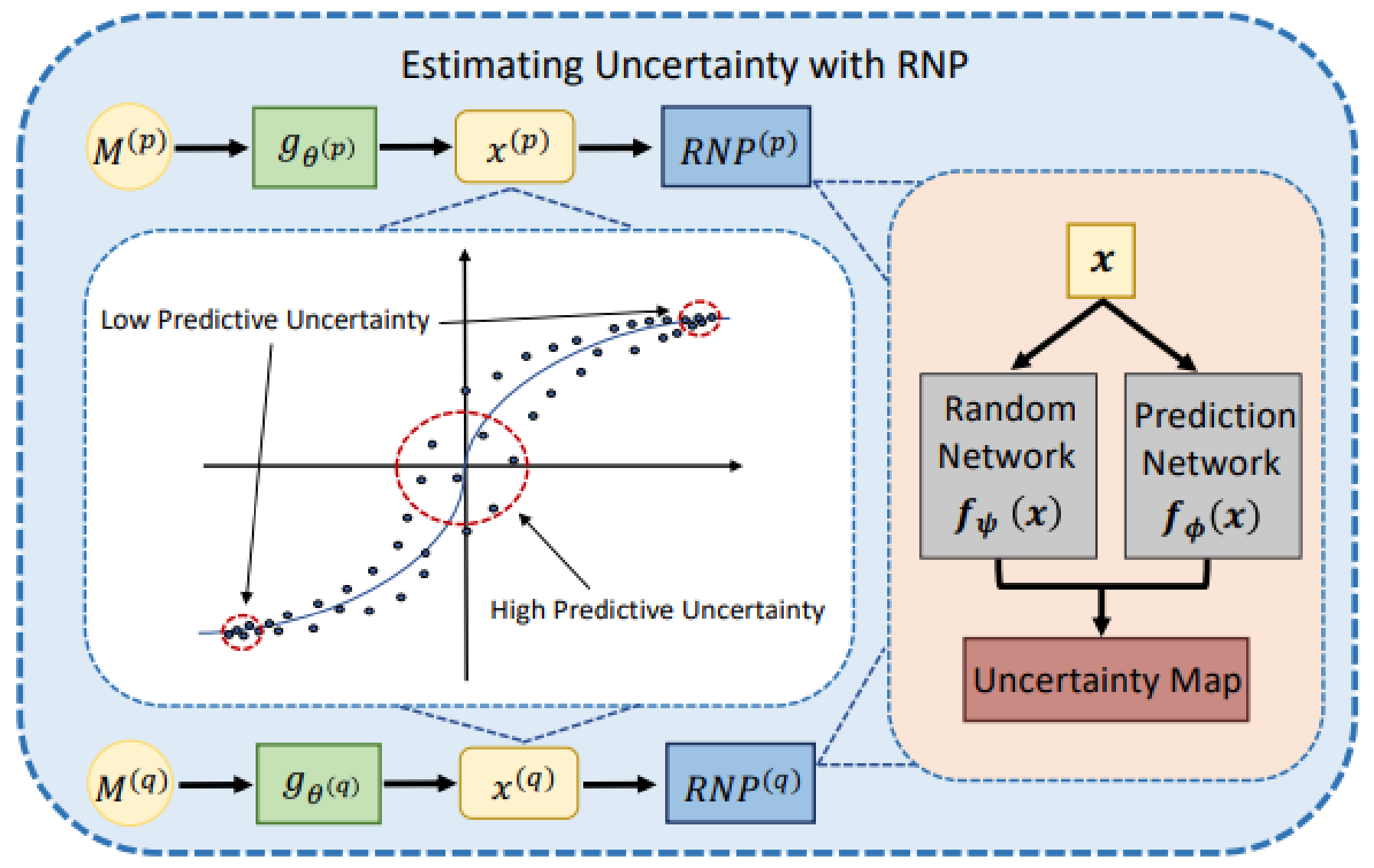
Estimating Uncertainty with RNP
$M^{(p)}$
$g_{\theta^{(p)}}$
$x^{(p)}$
$RNP^{(p)}$
Low Predictive Uncertainty
High Predictive Uncertainty
x
Random Network $f_\psi(x)$
Prediction Network $f_\phi(x)$
Uncertainty Map
$M^{(q)}$
$g_{\theta^{(q)}}$
$x^{(q)}$
$RNP^{(q)}$

Figure 1: The RNP Uncertainty Estimating Process.

The input data $M^{(p)}$ and $M^{(q)}$ are first processed by backbone models $g_{\theta^{(p)}}$ and $g_{\theta^{(q)}}$ that produce the features $x^{(p)}$ and $x^{(q)}$. Then the RNP modules have a fixed-weight random network $f_\psi(x)$ and a learnable prediction network $f_\phi(x)$ that tries to fit the output of the random network. The prediction network will fit better (i.e., with low predictive uncertainty) at more densely populated regions of the feature space, as shown in the graph. Hence, the difference between the outputs by $f_\psi(x)$ and $f_\phi(x)$ can be used to estimate uncertainty when processing a test input data.

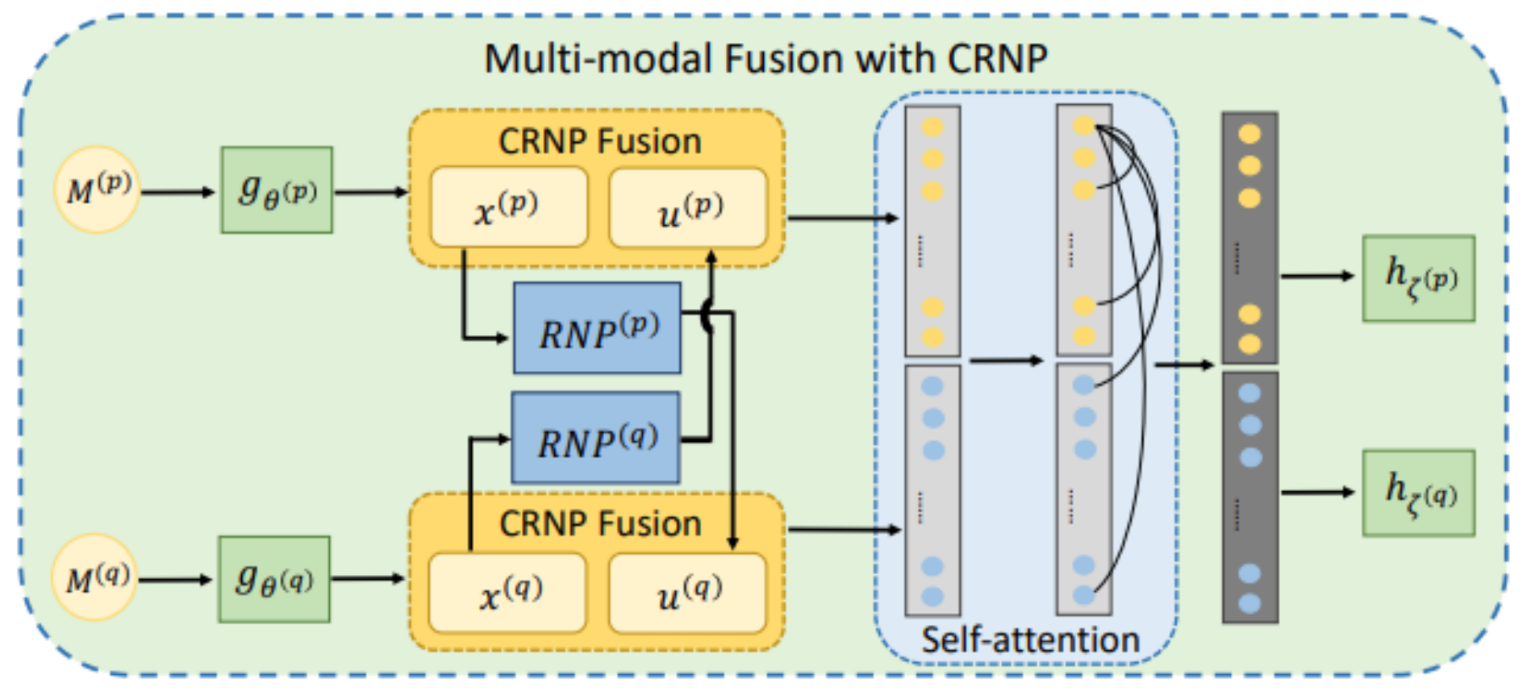

Figure 2: The overall framework of multi-modal fusion with our CRNP.

During the multi-modal fusion phase, the features of the two modalities $x^{(p)}$ and $x^{(q)}$ are cross-attended by the uncertainty maps produced by the RNP module from both modalities.

Results 3 & Visualization

Table 3: The performance of different models on computer vision classification datasets. The best results for each row are in bold.

Data	Metric	MCDO [11]	DE [21]	UA [15]	EDL [32]	TMC [14]	CRNP
Handwritten	Acc	0.9737	0.9830	0.9745	0.9767	0.9851	**0.9925**
	AUROC	0.9970	0.9979	0.9967	0.9983	**0.9997**	0.9996
CUB	Acc	0.8978	0.9019	0.8975	0.8950	0.9100	**0.9167**
	AUROC	0.9929	0.9877	0.9869	0.9871	0.9906	**0.9961**
Scene15	Acc	0.5296	0.3912	0.4120	0.4641	0.6774	**0.7057**
	AUROC	0.9290	0.7464	0.8526	0.9141	0.9594	**0.9734**

We now show results that demonstrate the effectiveness of CRNP on multiple CV classification tasks. The evaluation metrics include accuracy and multi-class AUROC on Handwritten, CUB and Scene15 datasets. Following Han et al., the comparison models include multiple uncertainty-aware models: Monte Carlo dropout (MCDO) that adopts dropout at inference as a Bayesian approximator; deep ensemble (DE), which uses an ensemble strategy to reduce uncertainty; uncertainty-aware attention (UA) that creates uncertainty attention maps from a learned Gaussian distribution; evidential deep learning (EDL) that predicts an extra Dirichlet distribution for all logits based on evidence; and trusted multi-view classification (TMC), which is a multi-view version of EDL.

We also conduct a visualization experiment in the figure that shows the MMWHS segmentation visualization (Sub Fig.1), T-SNE visualization of in and out of distribution data points produced by the uncertainty maps from the RNP module on the CT images from MMWHS (Sub Fig.2), and the CRNP uncertainty heat-maps for BraTS2020 images (Sub Fig.3).

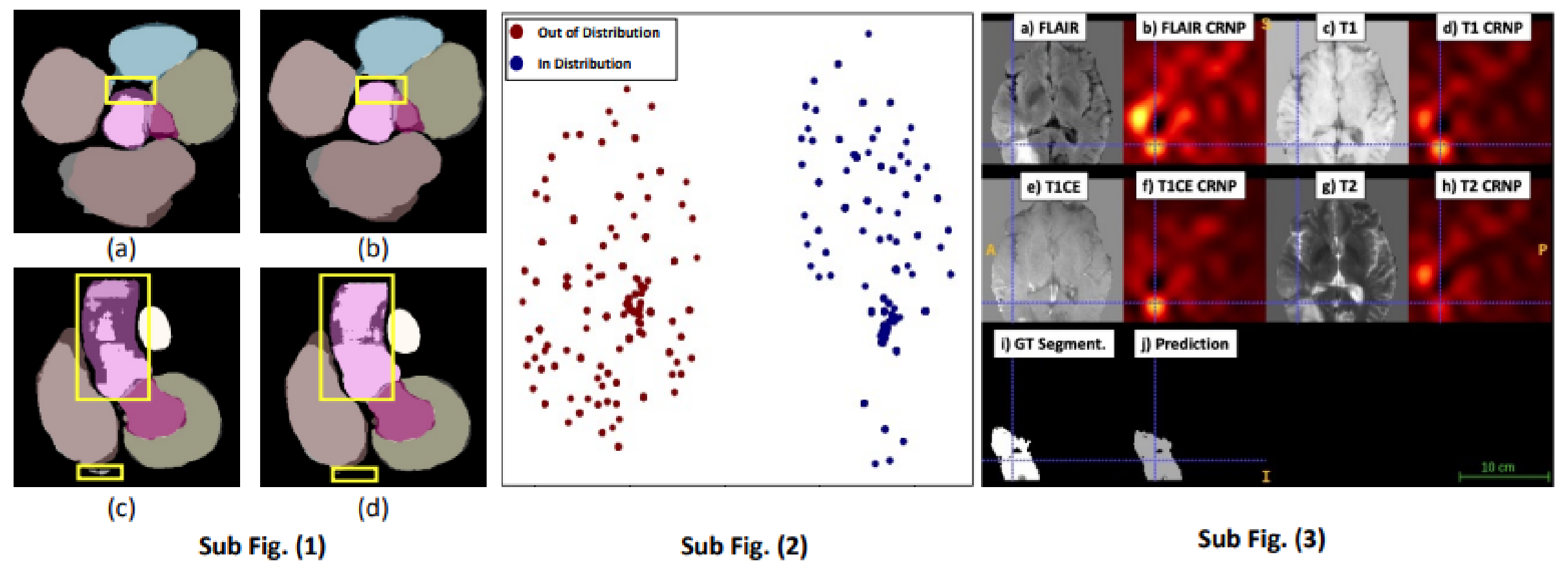

Figure 3: Visualization experiments of CRNP. Sub Fig.(1) shows a comparison between the segmentation of the proposed CRNP ((b) and (d)) and its Base model ((a) and (c)). Sub Fig.(2) shows the T-SNE graph of the in and out of distribution data points produced by the cross-modal RNP module. In the Sub Fig.(3), we show the CRNP uncertainty heat-maps.