Stochastic Segmentation Networks: Modelling Spatially Correlated Aleatoric Uncertainty

Imperial College London

project mira ERC StG 2017 erc BioMedIA

Miguel Monteiro, Loïc Le Folgoc, Daniel Coelho de Castro, Nick Pawlowski, Bernardo Marques, Konstantinos Kamnitsas, Mark van der Wilk, Ben Glocker

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

- Image segmentation is inherently ambiguous;
- Modelling this ambiguity enables predicting **multiple plausible segmentations for the same image**;
- Applications not only in medical imaging but also in computer vision.

Even human experts disagree on the shape of different lung nodules:

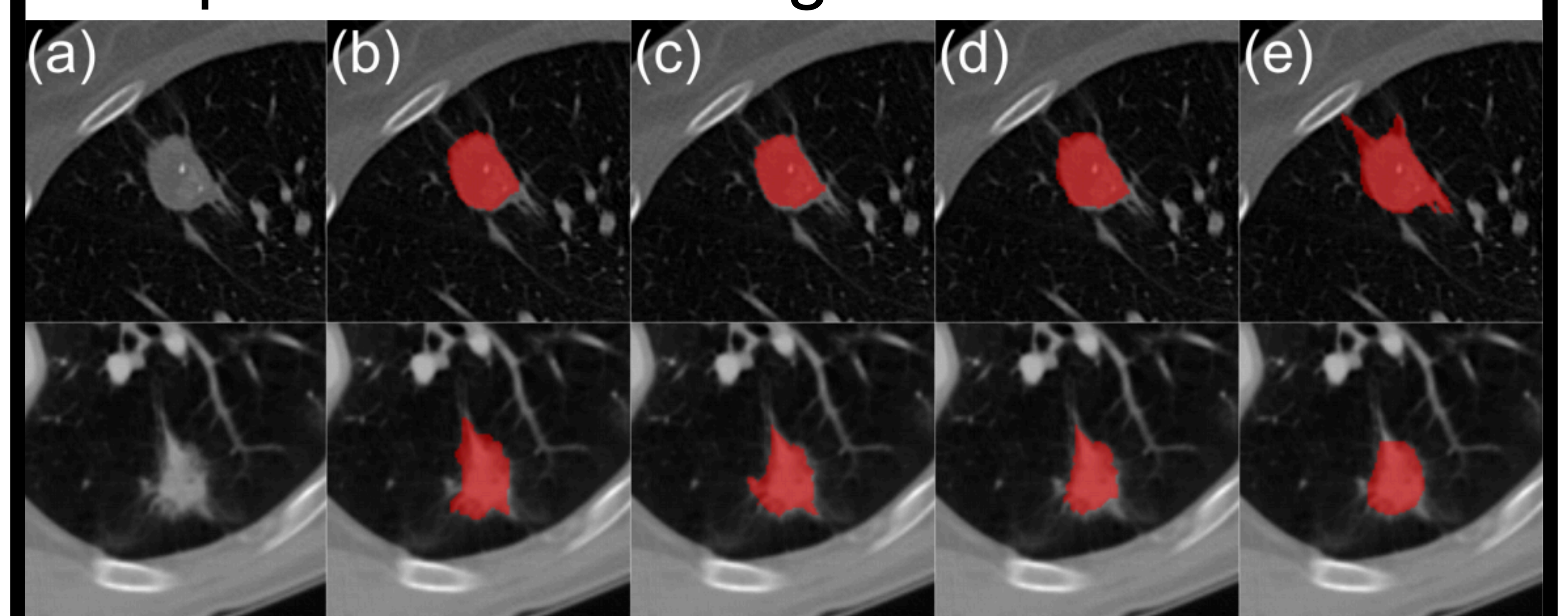

Image (a), ground-truth segmentation from four different human experts (b-e).

Problem: Sampling from the posteriors of a standard neural network produces spatially incoherent results because the pixels are treated independently:

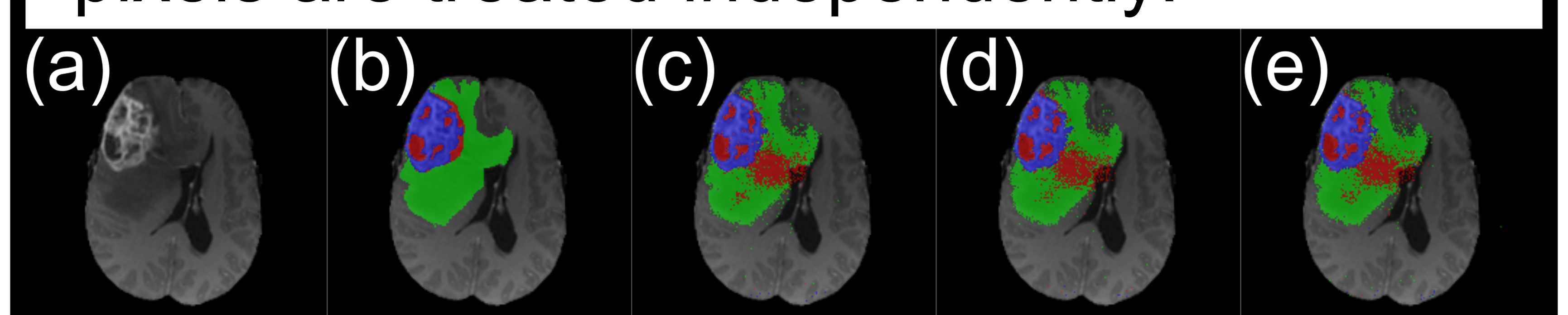

Image (a), ground-truth segmentation (b), three random samples (c-e). Note these are 2D slices of 3D images.

Methods

Solution: To make the logits dependent of each other by modelling them with a low-rank multivariate normal distribution:

$$p(\boldsymbol{\eta} \quad \boldsymbol{x}) = \mathcal{N}(\boldsymbol{\eta} \quad \boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\Sigma}(\boldsymbol{x})),$$

where $\boldsymbol{x}$ is the image and $\boldsymbol{\eta}$ the logit-map.

- The non-diagonal covariance matrix, $\boldsymbol{\Sigma}$, models dependencies between pixels;
- Its low-rank parametrisation, $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{P}^T + \boldsymbol{D}$, enables the distribution to be efficiently computed by a neural network even for large 3D images;
- The overhead is only predicting the three maps that describe the distribution, instead of one, at the output of the network;
- The rank of the covariance matrix can be adjusted to control the expressiveness of the distribution;
- The new loss function can be obtained using Monte-Carlo integration:

$$-\log p(\boldsymbol{y} \quad \boldsymbol{x}) \approx -\log \frac{1}{M} \sum_{m=1}^{M} p(\boldsymbol{y} \quad \boldsymbol{\eta}^{(m)}),$$

$$\boldsymbol{\eta}^{(m)} \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\Sigma}(\boldsymbol{x})).$$

Results

Multiple spatially consistent samples for the same image:

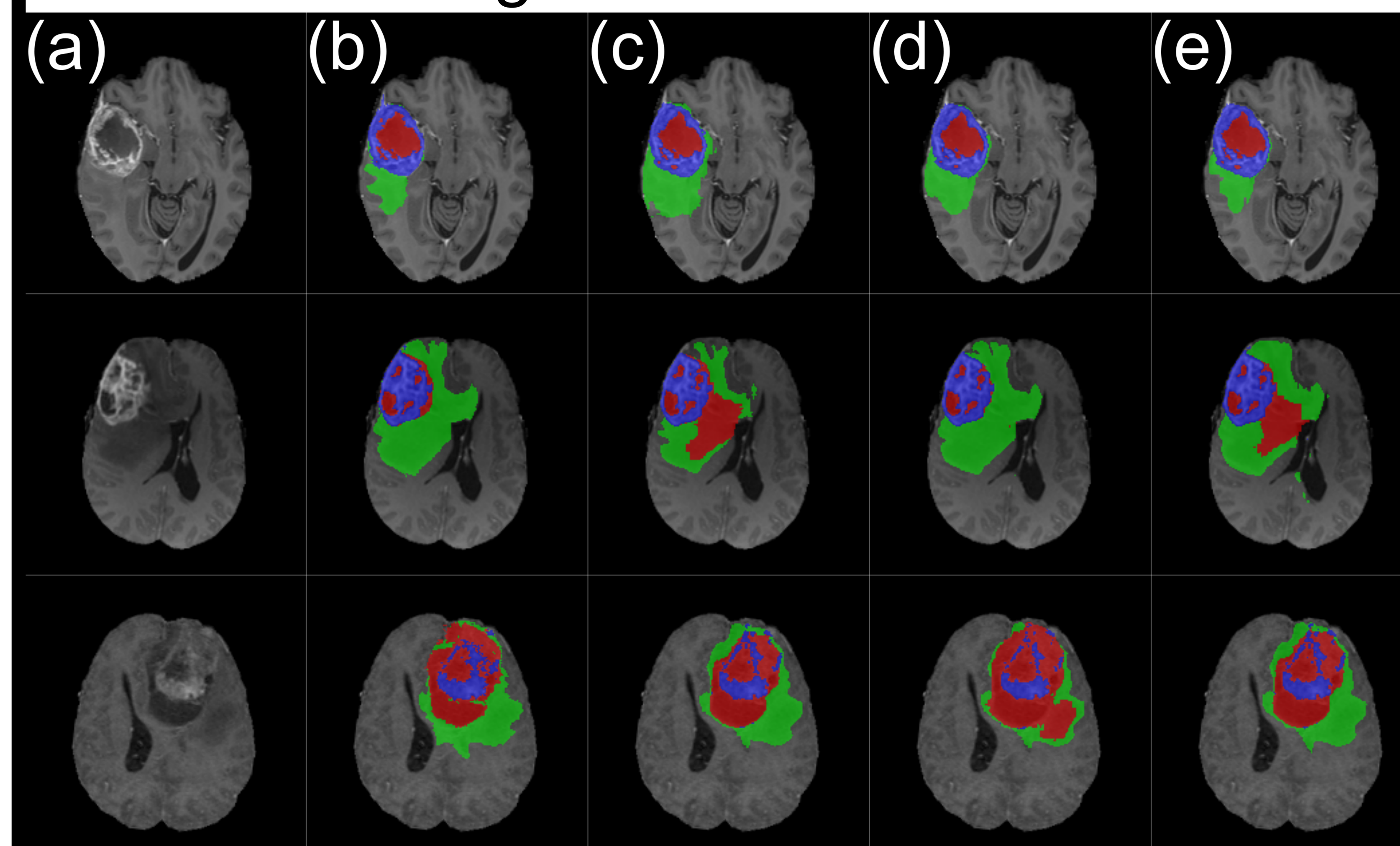

Image (a), ground-truth segmentation (b), three random samples (c-e). Note these are 2D slices of 3D images.

Comparison with SOTA:

- SOTA algorithms are based on conditional VAEs and strict, memory-intensive architectures;
- In contrast, our approach is light-weight and flexible enabling it to be used over any neural network architecture;
- We can generate infinitely many samples from one forward pass;
- In lung nodule segmentation, we obtain equal or better results in predictive performance and distance to the human expert distribution.