Appendix - Weakly Supervised Object Localisation (WSL)

In (Baumgartner et al., 2016), WSL was performed by exploiting the pixel-level saliency map obtained by guided-backpropagation, followed by ad-hoc procedure to extract bounding boxes. The same heuristics can be applied for the given network, however, owing to the attention map, we can device a much efficient way of performing object localisation. In particular, we generate object location by simply: (1) blur the attention maps, (2) threshold the low activations, (3) perform connected-component analysis, (4) select a component that overlaps at each scale and (5) apply bounding box around the selected components. In this heuristics, backpropagation is not required so it can be executed efficiently. We note, however, attention map outlines salient region used by the network to perform classification; in particular, it does not necessarily agree with the object of interest. This behaviour makes sense because some part of object will appear both in the class as well as background frame until the ideal plane is reached. Therefore, the quantitative result is shown in 7, however, the result is biased. We however define new metric called *Relative Correctness*, which is defined as 50% of maximum achievable IOU (due to bias). We see that in this metric, the method achieves very high results, indicating that it can detect relevant features of the object of interest in its proximity.

Table 7: WSL performance for the proposed strategy with AG-Sononet-16. Correctness (Cor.) is defined as $IOU > 0.5$. Relative Correctness (Rel.) is defined as $IOU > 0.5 \times \max(IOU_{class})$.

	IOU Mean (Std)	Cor. (%)	Rel. (%)
Brain (Cb.)	0.69 (0.11)	0.96	0.96
Brain (Tv.)	0.68 (0.12)	0.96	0.96
Profile	0.31 (0.08)	0.00	0.80
Lips	0.42 (0.18)	0.36	0.60
Abdominal	0.71 (0.10)	0.96	0.96
Kidneys	0.73 (0.13)	0.92	0.98
Femur	0.31 (0.11)	0.02	0.58
Spine (Cor.)	0.53 (0.13)	0.56	0.76
Spine (Sag.)	0.53 (0.11)	0.54	0.94
4CH	0.61 (0.14)	0.76	0.86
3VV	0.42 (0.14)	0.34	0.62
RVOT	0.56 (0.15)	0.70	0.76
LVOT	0.54 (0.15)	0.62	0.80