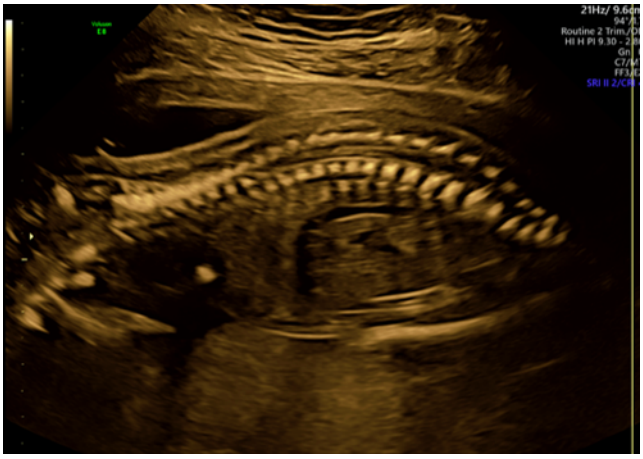
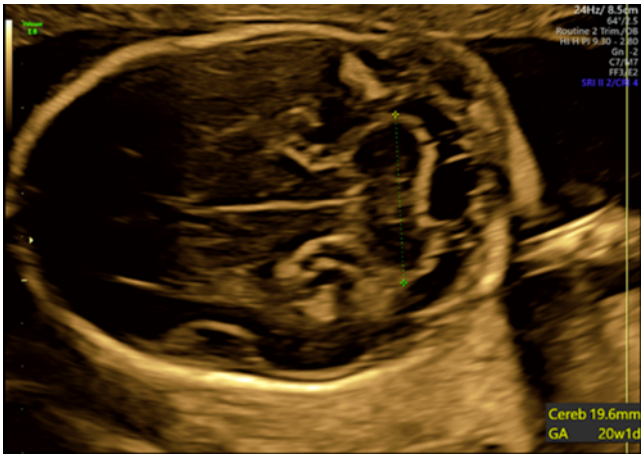
5 Supplementary Material



(a) Ground truth: **'we can see the cervical spine'**
Generated: **'the spine'**



66

(b) Ground truth: **'measuring the cerebellum'**
Generated: **'the cerebellum has a good measurement'**

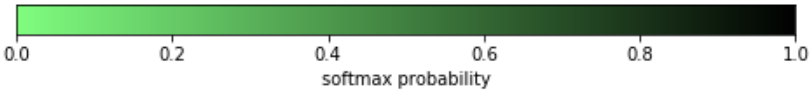
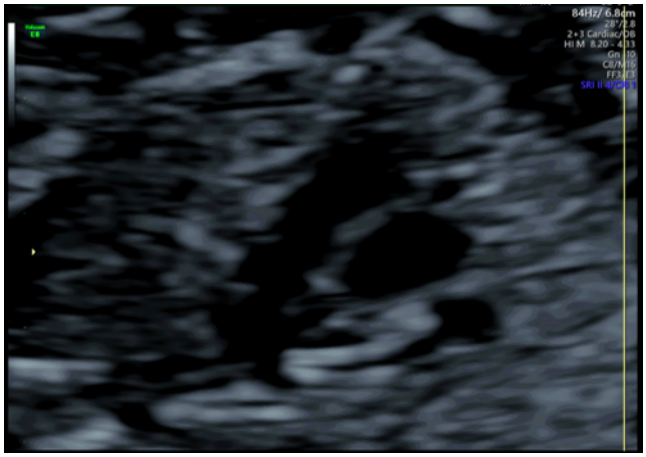
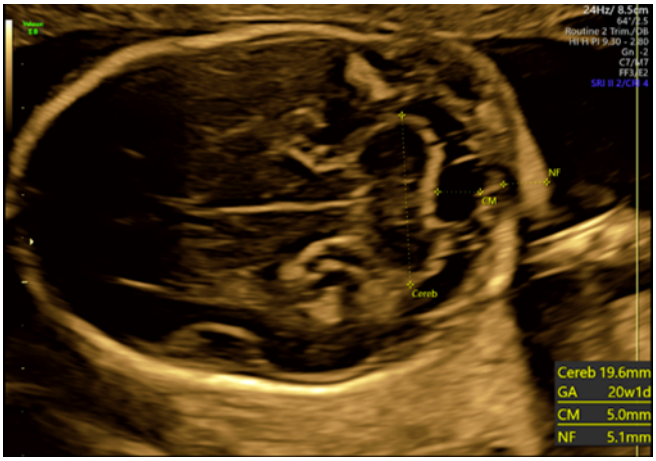


Fig. 5: Ground truth and good generated captions for a couple of images in the test set. The higher the softmax probability associated with a generated word, the darker the green color of that generated word.



(a) Ground truth: **'three vessel trachea view'**
Generated: **'this is the three vessel trachea view'**



(b) Ground truth: **'the measurements of the nuchal fold are good'**
Generated: **'this is the cisterna magna and nuchal fold'**

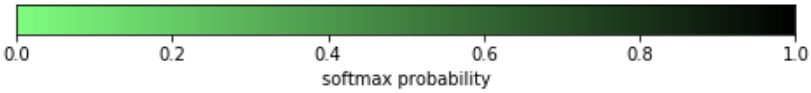
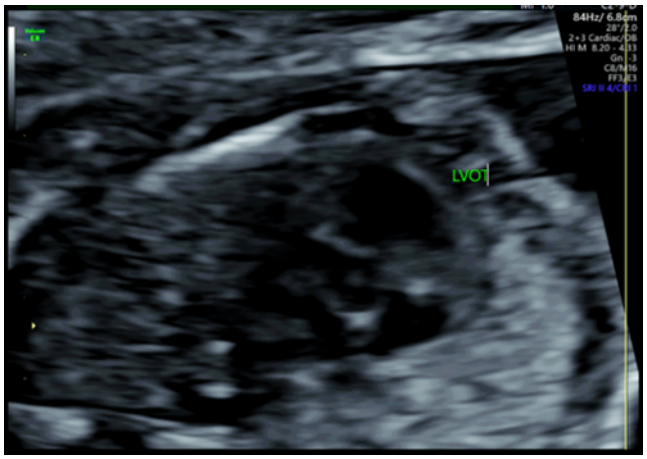


Fig. 6: Ground truth and good generated captions for another couple of images in the test set.



(a) Ground truth: **'this is the left ventricular outflow tract'**
Generated: **'the right ventricular outflow tract'**

(b) Ground truth: **'the aortic valve'**
Generated: **the right ventricular outflow tract'**

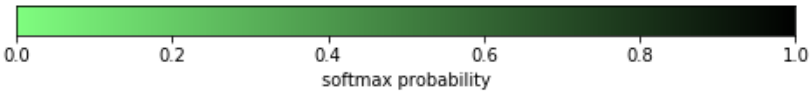
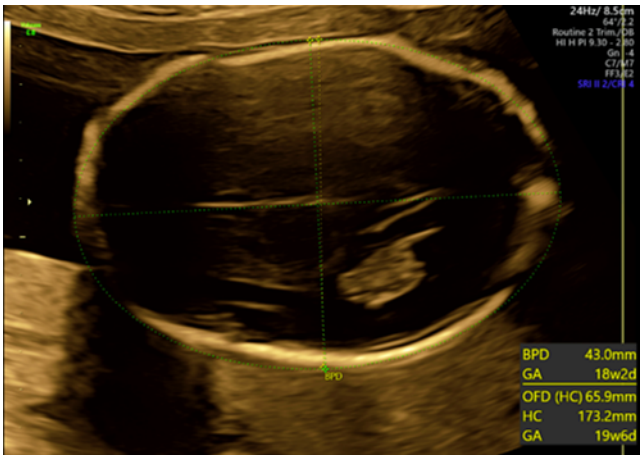


Fig. 7: Examples where the ground truth and generated captions do not match exactly, but the latter may describe the image contents with relevant terminology. This mismatch is reflected in the low objective scores in the Results section. Also, the confusion between heart views is evident in Fig. 7a. Please note that the words 'aortic' and 'valve' in the ground truth caption of Fig. 7b are not in the training set.



(a) Ground truth: **'now measuring the head side to side and around the head'**
Generated: **'this is the cisterna magna and nuchal fold'**

(b) Ground truth: **'rib'**
Generated: **'we can see the stomach'**

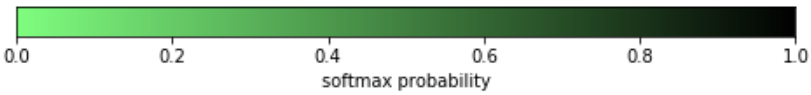


Fig. 8: Additional examples where the ground truth and generated captions do not match. Note that the stomach is visible in Fig. 8b, but the sonographer happened to be talking about a rib in this instance.