Supplementary Figure 2: Three cases where majority voting had a Likert Score less than 4. (Top) We believe that the top failed due to similarity in appearance to the gallbladder. (Middle, Bottom) Model over-segmented regions around the ablation zones.