**Supplemental Digital Content 1 Supplemental method**

This work uses a mask region-based convolutional neural network (Mask R-CNN) to detect the masks of the distal ETT end and tracheal bifurcation. The backbone of the neural network was composed of 50-layer ResNeXt networks (32×4d) with a recently proposed feature pyramid network. The head architecture of the neural network is extended by the Faster regional-based convolutional neural network (Faster R-CNN) heads.

In the inference process, only the objects with the maximal score for each class were retained. Subsequently, a post-processing procedure developed based on the preliminary evaluation of the algorithm performance was performed. This rule-based feature extract method was applied to identify the exact feature points. The bounding boxes and the masks detected by the mask region-based convolutional neural network were used to identify candidate feature points of the ETT tip and carina.



The exact feature points of the ETT tip and carina can be obtained as follows. In the masking processes, skeletonization was used to identify the skeleton of the masks of the distal ETT end and the trachea bifurcation. For the ETT tip, the feature point was determined by the lowest point in the skeleton of the distal ETT end mask. To find the feature point of the carina, a slide window with a size of 15×15 was applied to identify the central point in the skeleton of the trachea bifurcation mask. In addition, the Canny edge detection was used to detect the edge of the trachea bifurcation mask. According to the central point, a patch of the mask edge with a size of 100×150 was cropped. Subsequently, another slide window with a size of 7×7 was applied to identify the feature point (i.e., the carina) in the patch.

For the location of the ETT tip, if the bounding box was identified, the central point at the bounding box was the exact feature point of the ETT tip. However, because the bounding box was considerably small, bounding boxes could not be detected in about 10% of the images. In these cases, the mask of the distal ETT end was used to identify the exact feature point of the ETT tip.

 For the location of the carina, the bounding box results had a low mean error and high standard deviation; on the contrary, the mask results had a high mean error and low standard deviation. Therefore, we decided to use the mask results to eliminate the worst cases of the bounding box results. If the bounding box of the carina did exist and the distance between the center of the bounding box and the feature point obtained by the mask was not greater than 100 pixels (0.139 mm), the central point at the bounding box was the exact feature point of the carina. However, if the bounding box existed and the distance between the center of the bounding box and the feature point obtained by the mask was greater than 100 pixels, or the bounding box was not detected, the mask result was preferred as the exact feature point of the carina.