Assessment of Endotracheal Tube Position Relative to the Carina Using A Cascaded Convolutional Neural Network Approach

Su Kara[1*], Jake Y. Akers[2], and Peter D. Chang[2]

[1]*Capistrano Valley High School, Mission Viejo, CA 92692, USA*
[2]*University of California, Irvine, CA 92697, USA*

Abstract.—Rapid and accurate assessment of ETT location via a deep learning system may significantly expedite patient care in the ICU setting, where a misplaced ETT can result in severe patient morbidity and mortality. We propose a deep learning-based algorithm for characterization of ETT position relative to the carina from chest radiographs. Using the MIMIC-CXR dataset, an open-source critical care database, a natural language processing technique based on regular expressions was used to parse 227,835 patient reports (corresponding to 377,110 images). A subset of 16,000 reports were identified with a high probability of either presence or absence of ETT (8,000 reports in each category). Three different CNN algorithms were created. The first algorithm comprised of a regression network designed to output the estimated coordinate of the carina. Using this input, a second CNN binary classifier network was used to predict presence or absence of ETT. Finally, if an ETT was detected, a third CNN regression network was designed to output the estimated coordinate of the distal ETT tube. Upon five-fold cross validation, carina coordinate location was estimated within 0.66cm of ground-truth annotations. Classification accuracy of presence or absence of ETT demonstrated an accuracy of 96.07%. ETT tip position was predicted within 0.63cm of annotations. Final prediction of distance from the ETT tip to carina was estimated within 1.09cm of ground-truth measurements in the MIMIC-CXR patient reports. A serial multi-step CNN approach demonstrates high accuracy for localization and assessment of ETT position relative to carina.

Introduction

An endotracheal tube (ETT) is a plastic tube placed through the mouth into the trachea to help a patient breathe (Eldridge and Doru 2020) as shown in Fig. 1. ETT is connected to a ventilator to deliver oxygen to the lungs. Common practice is to secure ETT at 23cm for men and 21cm for women. The desired position of an ETT is 5±2cm above the carina (Peitzman et al. 2019).

[*]Corresponding author: su.kara@gmail.com

An ETT misplacement may cause serious complications such as the collapse or hyperinflation of a lung (Gupta et al. 2014). ETT position is usually assessed on a frontal chest radiograph. A prompt and precise measurement of the ETT tip position relative to the carina from a chest radiograph is a critical need in the acute care setting. In this study, we propose a deep learning-based algorithm for characterization of ETT position relative to the carina from chest radiographs.

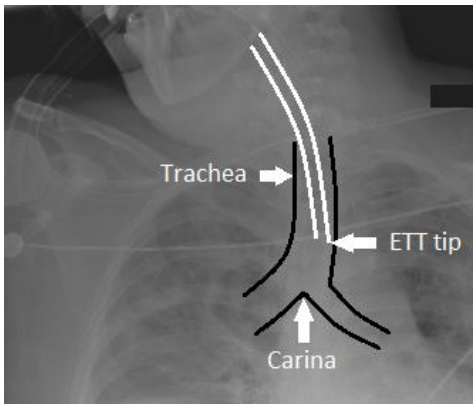

Fig. 1. Trachea, ETT tip, and carina.

Specifically, we separate the task into three components reflective of the interpretative process of a human radiologist: localization of the carina; determination of the presence or absence of an ETT; localization of the ETT tip and measuring the distance to the carina. Each individual component of this algorithm was implemented via a convolutional neural network (CNN). A serial cascade of deep learning CNNs can be used to determine the distance of an ETT from the carina with high accuracy.

Materials and Methods

For all development purposes throughout this study, a Google Colaboratory (Colab) Pro account was used with a Python 3 runtime type and a GPU hardware accelerator. TensorFlow 1.15.0, an open-source software library, was used to develop the CNN algorithms.

The MIMIC Chest X-ray (MIMIC-CXR) Database v2.0.0, an open-source critical care dataset of chest radiographs in DICOM format with free-text radiology reports, served as the main data source of this research project (Goldberger at al. 2003; Johnson et al. 2019ab). The dataset contained 377,110 images corresponding to 227,835 reports of 65,389 patients performed at the Beth Israel Deaconess Medical Center in Boston, MA, between 2011 and 2016.

A natural language processing technique based on regular expressions was used to parse all patient reports in the MIMIC-CXR dataset. The following regular expression was implemented in Python to identify the reports that contained endotracheal or ETT with a distance to carina: "(\bendotracheal\b|\bett\b)+\D+(\d*\.*\d*\-*\d*\.*\d*)\s*cm\D+\bcarina[l]*\b".

From the initial cohort of 227,835 studies, a subset of 8,000 patient reports were identified with the presence of ETT and another set of 8,000 reports represented the absence of ETT with a high probability. These 16,000 studies corresponded to 24,010 chest radiographs in the dataset.

Chest radiograph images were in Digital Imaging and Communications in Medicine (DICOM) format. DICOM is a common format for medical images, which contains structured meta-data

associated with one or more images. The identified 24,010 images were downloaded and converted from DICOM to Numpy array format for image processing in a CNN.

Images were converted to a uniform size of 512x512 during the preprocessing stage. Since the gold standard of detecting the position of ETT relative to carina required frontal images, only the frontal projection radiographs were isolated to have a total of 17,050 images (8,848 ETT positive and 8,202 ETT negative).

A CNN architecture was created to train with the annotated images by using a regression algorithm (Fig. 2). The custom algorithm was composed of simple VGG-like blocks (convolution to batch normalization to ReLU to max pool) implemented with 5x5 convolutional filters. A total of 5 convolutional blocks were used, with gradually increasing channel depth from 16 to 64. The final feature map was then flattened and connected to a hidden layer of size 128. The regression algorithm to estimate the coordinate of the carina was trained to optimize an L2 distance between the final logit score and the expected normalized coordinate point.

Training was initialized from random weights. Optimization was implemented via the Adam method with a learning rate of 0.001 and a batch size of 64. Approximately 10,000 iterations were required for algorithm convergence. A five-fold cross validation technique was used to estimate performance.

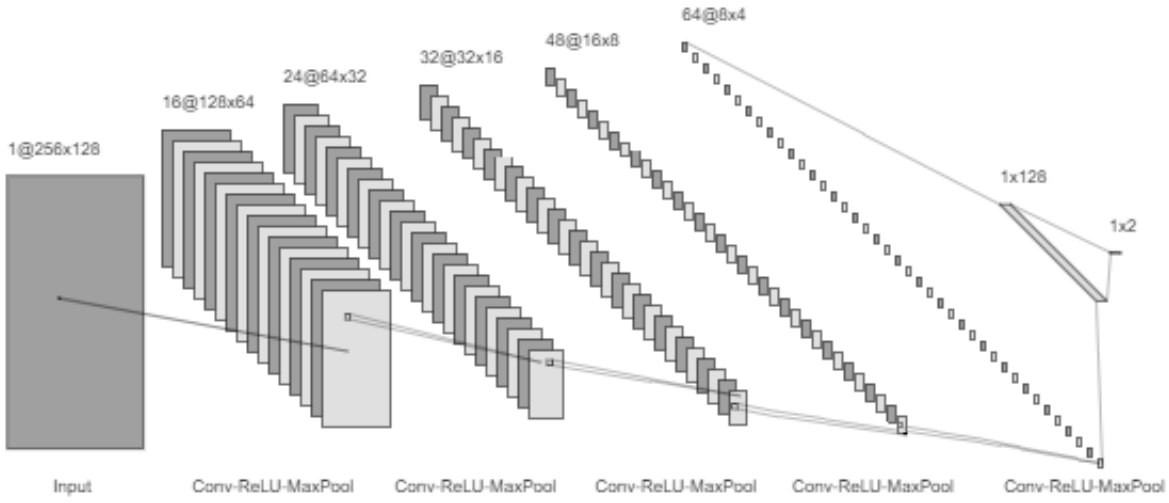

Fig. 2. Architecture diagram of common CNN backbone. All three architectures shared an identical backbone design with 5 convolutional layers followed by a hidden layer.

A subset of the frontal images was converted to grayscale PNG format and resized from 512x512 to 1024x1024 to easily annotate the (y, x) coordinate of the carina. From this set of images, carina location was manually annotated on 4,000 images.

A CNN algorithm was used to estimate the position of carina on all 512x512 frontal images. A custom algorithm was developed to crop a 256x128 portion of each image around the carina to have the position of carina located at a proportion of (0.75, 0.5) as shown in Fig. 3.

The newly generated 17,050 cropped images were used to train a CNN binary classification algorithm, where their existence in the list of reports created by the regular expression served as Boolean labels. The classifier algorithm was trained to optimize a standard binary software cross entropy loss. The CNN algorithm was used to classify images as ETT positive or negative.

A subset of the cropped images was converted to grayscale PNG format and resized from 256x128 to 512x256 to annotate the (y, x) coordinate of the ETT tip. On the selected images, ETT tip location was annotated on 3,000 images by manual inspection.

A new CNN regression algorithm was created to locate the position of ETT tip on the 3,000 annotated images. The CNN algorithm was used to estimate the position of ETT tip on all the cropped images. Those coordinates were later used to calculate the distance from the ETT tip to the carina that was located at (191, 63) on each cropped image.

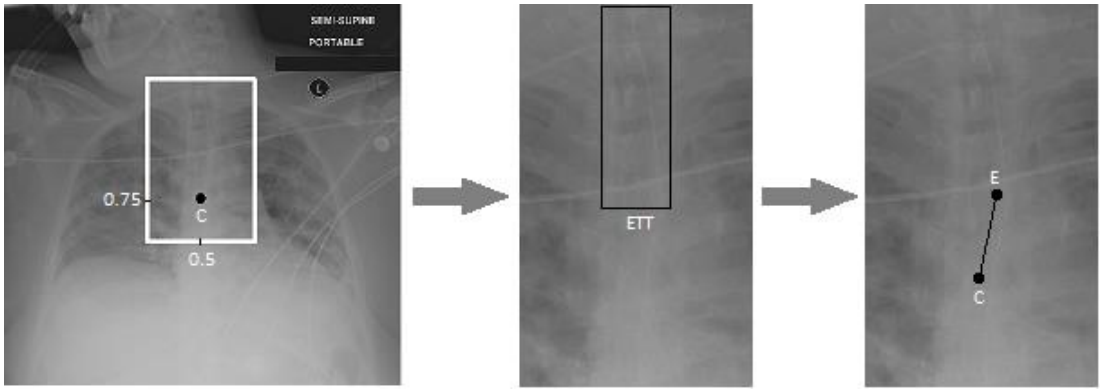

Fig. 3. Serial three-step CNN architecture. In the first step, the CNN derived estimate for carina location (C) is used to create a 256 x 128 bounding crop of the upper airway. In the second step, the presence or absence of ETT is determined via a CNN classifier. In the third step, the final ETT tip coordinate (E) is estimated and used to calculate the distance relative to the carina.

A total of three different CNN algorithms were created (Fig. 3). The first algorithm comprised of a regression network designed to output the estimated (y, x) coordinate of the carina. From this information, a 256 x 128 crop was generated of the upper airway. Using this input, a second CNN binary classifier network was used to predict presence or absence of ETT. Finally, if an ETT was detected, a third CNN regression network was designed to output the estimated (y, x) coordinate of the distal ETT tube.

Results

During the preprocessing phase, the image dimensions decreased to one-sixth of their original values. Since the pixel spacing on the original images was 0.0138cm on average, the pixel width and height on the new images became six times greater. Distance between the coordinates of each ground-truth annotation and the coordinates of the corresponding CNN estimated position was first calculated in terms of pixels by using the Pythagorean theorem. Later that number was converted from pixels to centimeters (cm) by multiplying the distance in pixels by 0.0828, i.e. 6 times 0.0138.

A five-fold cross validation technique was used to train and test the first CNN regression algorithm with the 4,000 annotated images in 90 minutes. The algorithm detected the position of carina within 0.66cm of ground-truth annotations (Fig. 4).

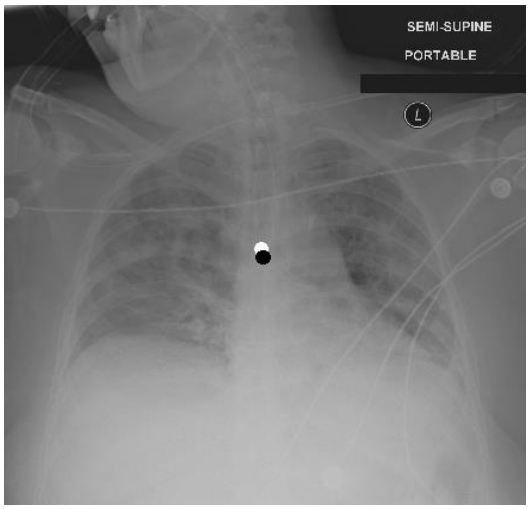

Fig. 4. Detection of carina position. Manually annotated carina position is shown as the white circle, while the CNN derived estimate for its location is shown as the black circle at the center.

The model generated by the first CNN algorithm was used to detect the position of carina and to crop around it. The second CNN algorithm was trained on these 17,050 cropped images to classify them as ETT-positive or ETT-negative. This goal was achieved by labeling every image as 0 or 1 based on whether the study associated with that image existed in the list of reports created by the regular expression. A five-fold cross validation run was completed in 20 minutes with a high accuracy of 96.07% (Table 1).

Table 1. Performance statistics of the binary classification CNN algorithm to detect the presence or absence of ETT on the 17,050 cropped images.

Statistic	Value	95% Confidence Interval
Sensitivity	96.41%	96.00% to 96.78%
Specificity	95.71%	95.25% to 96.14%
Disease Prevalence	51.89%	51.14% to 52.65%
Positive Predictive Value*	96.04%	95.63% to 96.41%
Negative Predictive Value*	96.11%	95.68% to 96.49%
Accuracy*	96.07%	95.77% to 96.36%

* Dependent on disease prevalence

The third CNN algorithm was a regression algorithm to find the position of ETT tip on the cropped images. A five-fold cross validation was used to train and test the algorithm with the 3,000 annotated images in 10 minutes. The CNN algorithm detected the position of ETT tip within 0.63cm of ground-truth annotations (Fig. 5).

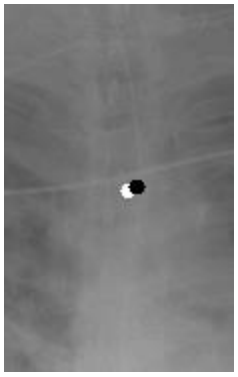

Fig. 5. Detection of ETT tip position. Manually annotated ETT tip position is shown as the white circle, while the CNN derived estimate for its location is shown as the black circle.

The model generated by the third CNN algorithm was used to detect the position of the tip of ETT on all the ETT-positive images. Each patient report associated with each of those images had a value for the distance between the carina and the ETT tip. That value was treated as the ground-truth distance to check against. The model was run on each image to predict the position of ETT tip. Since the carina was always located at (192, 128), the distance between the predicted position of ETT tip and the location of carina was calculated separately. These distance measurements were estimated within 1.09cm of ground-truth measurements in the MIMIC-CXR patient reports.

Discussion

A serial multi-step CNN approach implemented via a custom VGG derived architecture, trained on 17,050 patient images, demonstrates high accuracy for localization and assessment of ETT position. Compared to a single binary classifier, the proposed architecture is able to provide explicit feedback regarding the exact position of the ETT relative to the carina, helping guide clinical decision making. By decomposing the problem into small tasks, the final trained CNN was able to be implemented with a relatively small parameter footprint.

To achieve the highest accuracy in the CNN binary classification algorithm, and to minimize the errors in the two CNN regression algorithms, several parameters were changed in the common architecture and the training procedures. Number of layers in the architecture was arranged from three to six layers, and the current number of five layers provided the best performance for all algorithms. Likewise, different filter depths between 32 and 256 were tested, and 64 achieved the best results. On the training side, number of iterations was changed in the range of 2,000 and 20,000, and 10,000 was the top performer without overfitting the training data. Batch size was also tested with several values between 8 and 128, and 64 was the best setting.

MIMIC-CXR database offers a vast amount of data in terms of patient reports and images. However, there isn't any labeling of those images to make them readily available for processing in a CNN. To overcome this critical limitation, natural language processing techniques are required to turn reports into labels. Regular expressions were utilized to parse all the reports and match the related keywords and patterns.

Another limitation of this study is the lack of pixel spacing information for some of the DICOM images in the MIMIC-CXR database. This may have affected the conversion accuracy from pixels to centimeters in the regression CNN algorithms as the distance in pixels was multiplied by the same constant value of 0.0828 for all images in the test dataset.

All images of a patient were either used for training or validation, but not both. This was because the two images were likely to be similar and not to be mixed. If two images from the same patient were used in training and validation, the algorithm could memorize them, and our estimate of accuracy would be erroneously high. Since patient images were kept in 10 main directories (p10-p19), for each iteration of the five-fold cross validation, 8 directories were used for training and 2 directories were spared for testing to ensure a clear separation.

Cropping images around the carina improved both accuracy and performance. Even though the first and third CNN algorithms shared the same architectural backbone and similar number of images to train on, the first one took nine times longer to process. This could be attributed to the image size of 512x512 used in the first algorithm being eight times the cropped image size of 256x128 used in the third one.

The CNN binary classification algorithm detects the presence or absence of ETT with a very high accuracy of 96.07% because it was trained with 17,050 cropped images. The first CNN regression algorithm was able to locate the position of carina within 0.66cm as it was trained with only 4,000 annotated images and they were of size 512x512. The third CNN had a similar performance to find the ETT tip position within 0.63cm even though the 3,000 cropped images were much smaller. Training for the carina was an easier task as there was a single point on the image. However, the ETT tip was sometimes on the left corner, sometimes on the right, and sometimes at the center of the ending line. Since there was no single point to train on, there was always a small distance between an annotation and a prediction.

The two CNN regression algorithms give the impression of performing worse than the classification algorithm, but it's hard to compare the two. Regression algorithms aim to find the coordinates of one pixel on an image with several thousands of them, while the classification algorithm only tries to recognize an object without any precision.

Google Colab is an extremely valuable tool that lets researchers run their CNN algorithms in TensorFlow for free. After running several algorithms with a GPU hardware accelerator for a few days, it stops providing the accelerator option. GPU accelerator is so powerful that it completes the CNN training process ten times faster than a regular processor with no accelerator. Therefore, a Google Colab Pro account had to be purchased at a reasonable price to avoid the limitations of the free account and to test the CNN algorithms with several different configuration options.

A misplaced ETT paired with delays in physician notification can cause severe complications and damages to a patient's health. Deep learning systems may significantly improve patient care in the ICU environment by locating the position of ETT and its distance to carina in a matter of seconds with high accuracy.

This study can be enhanced with development of an application with a user interface, which takes a frontal chest radiograph as input, preprocesses it, runs the three CNN algorithms, and displays the position of ETT tip relative to carina and the distance between them. Full automation of this process will minimize errors by preventing human intervention and will improve patient health with fast response times.

Acknowledgments

We would like to thank Scott Refugio and Charles Lin of University of California, Irvine, for letting us use their viewer/coordinate annotation tool that helped us annotate the position of carina and ETT tip on thousands of images. We also thank Gloria Takahashi and Dr. Kimo Morris for their help throughout the Research Training Program of Southern California Academy of Sciences.

Literature Cited

Eldridge, L. and P. Doru. 2020. How an endotracheal tube is used: understanding the purpose, procedure, and possible risks. Verywell Health. https://www.verywellhealth.com/endotracheal-tube-information-2249093

Goldberger A.L., L.A.N Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C-K Peng, and H.E. Stanley. 2003. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 101(23):e215-e220.

Gupta, P.K., K. Gupta, M. Jain, and T. Garg. 2014. Postprocedural chest radiograph: Impact on the management in critical care unit. Anesthesia, essays and researches, 8(2):139–144. https://doi.org/10.4103/0259-1162.134481

Johnson, A.E.W., T.J. Pollard, S.J. Berkowitz et al. 2019a. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Sci Data. 6(317). https://doi.org/10.1038/s41597-019-0322-0

———, ———, ———, R. Mark, and S. Horng. 2019b. MIMIC-CXR Database (version 2.0.0). PhysioNet. https://doi.org/10.13026/C2JT1Q

Peitzman, A.B., D.M. Yealy, T.C. Fabian, and C.W. Schwab. 2019. The trauma manual: trauma and acute care surgery. Pp. 22-33 in Airway Management and Anesthesia. Philadelphia: Wolters Kluwer. 1040 pp.