Cervical Spine Fracture Detection and Localization

Naman Raghuvanshi, Siddharth Rayabharam and Sumant Suryawanshi

Contributing authors: nvr5386@psu.edu; nqr5356@psu.edu; szs7220.psu.edu;

Abstract

Purpose: To develop a deep learning model for the detection and localization of cervical spine fractures in the axial CT scans. **Methods:** We use the dataset consisting of cervical spine CT scans provided by the Radiological Society of North America (RSNA). The dataset consists of 3000 studies of individual patients for the training and testing combined. Out of these 3000 patient studies, 83 of these studies also contains segmentation data. Additionally, for x studies the dataset contains bounding box coordinates data. The dataset population is split into 90% for training and 10% for validation. We use EfficientNetV2[1] to learn the segmentation, YOLOv5 model for detecting the fractures and drawing bounding boxes around the fracture area. **Results:** The segmentation model predicts the vertebrae label with the accuracy of 95.13% whereas the fracture detection model predicts fractures with an accuracy of 94%. The fracture detection model also draws bounding boxes localizing the fractures and provides the objectiveness of the prediction. **Conclusion:** We have demonstrated that deep learning models are able to provide good fracture detection predictions and localization from the CT radiographs. We have shown that the CNNs holds promise at both worklist prioritization and assisting radiologists in cervical spine fracture detection on CT radiographs.

Keywords: Cervical Spine, Fracture Detection, Deep Learning

1 Introduction

Cervical spine injury is very common injury with more than 3 million cases per year that are being evaluated for cervical spine injury in North America[2]. In United States, more than 1 million patients with blunt force injury are suspected to suffer cervical spine injury[3]. Since cervical spine injury is associated with high morbidity and mortality, quick diagnosis of the injury is crucial. Any delay in diagnosis may result in devastating consequences for the patient. So, any additional aid to the radiologists can reduce the morbidity or mortality of the patient.

In recent years, a machine deep learning technique known as deep convolutional neural network (DCNN) has been applied to image recognition tasks. DCNN's are well suited for images. So, they have been used extensively in the field of medicine to classify medical images.

In past few years, there have been many studies that have tried to use DCNN[4][5][6] on medical radiographs. In these studies, the reference standard for the training and testing images was based on the assessment of human readers determining which were visible, only within a radiograph. Many radiologist fail to detect "occult fracture" because of the difficulty in detecting such fracture in a radiograph. These extraction methodologies could adversely influence the classification accuracy and occult fracture being assessed as a "non-fracture case". A proficient algorithm may help identify and triage studies for the radiologist to review more urgently, helping to ensure faster diagnoses.

The purpose of our study is to develop an automated deep learning system for detecting cervical spine fractures using CT a gold standard annotated by radiologists, and to evaluate the diagnostic performance inclusive of the experienced readers in detecting cervical spine fractures on radiographs.

2 Literature Review

Many studies have been conducted to utilize deep learning techniques in telemedicine field. We have adequate amount of literature that is relevant to our project. However, The CNN model proposed by J.E. Small et al. to detect the cervical spine fracture detection with the accuracy of 92%[7]. According to their study, the radiologists are 95% accurate in the cervical spine fracture detection. The fractures missed by the proposed model and the radiologists were similar by level and location which is mostly included fractures which are obscured by CT beam attenuations. Yee Liang Thian et al.[8] have implemented Inception-ResNet Faster R-CNN architecture to detect and localize the fractures on wrist radiographs. The model detected and correctly localized 310 (91.2%) of 340 and 236 (96.3%) of 245 of all radius and ulna fractures on the frontal and lateral views, respectively.

Yoga Dwi Pranata et al. have implemented ResNet and VGG for automated classification and detection of calcaneus fracture in CT images[9]. The bone fracture detection algorithm incorporated fracture area matching using the speeded-up robust features (SURF) method, Canny edge detection, and

contour tracing. Their results showed that ResNet was comparable in accuracy, of 98%, to the VFF network for bone fracture classification but achieved better performance. Tsubasa Mawatari et al. have studied about the effect of deep convolutional networks on radiologists performance in the detection of hip fractures on digital pelvic radiograph. According to the study, the average AUC of the 7 readers was 0.832. The AUC of DCNN alone was 0.905. The average AUC of the 7 readers with DCNN outputs was 0.876. The AUC of readers with DCNN output were higher than those without[10].

3 Methodology

3.1 Dataset and Study Population

We use the dataset consisting of cervical spine CT scans provided by the Radiological Society of North America (RSNA)[11]. The dataset we are using is made up of roughly 3000 CT studies, from twelve locations and across six continents. Spine radiology specialists have provided annotations to indicate the presence, vertebral level and location of any cervical spine fractures. Each radiology study consists of multiple dcm files. A dcm file follows the Digital Imaging and Communications in Medicine (DICOM) format. It is the standard format used for storing medical images and related metadata. It dates back to 1983, although it has been revised many times. Out of these 3000 patient studies, 83 of these studies also contains segmentation data. Additionally, for x studies the dataset contains bounding box coordinates data. The dataset population is split into 90% for training and 10% for validation. The distribution of fractured and non-fractured dataset in the dataset is shown in the below two graphs.
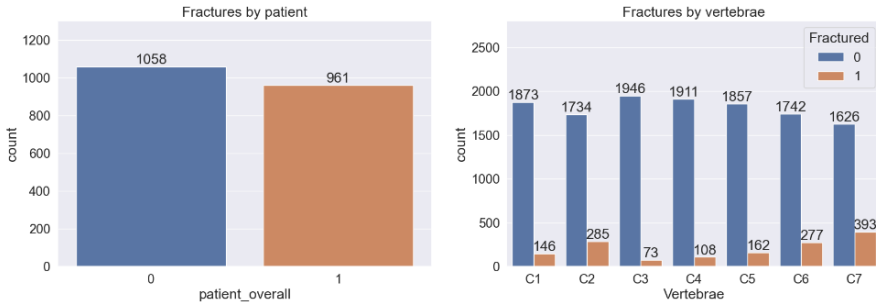

Fig. 1 a) Number of patients with fractures (blue) and without fractures (red). b) Fracture distribution among the vertebrae C1, C2, C3, C4, C5, C6 and C7.

3.2 Data Pre-processing

3.2.1 Windowing

Every CT radiograph contains details of multiple tissues. To highlight only certain anatomical features of the radiograph we use *windowing*. Windowing, also known as grey-level mapping, contrast stretching, histogram modification or contrast enhancement is the process in which the CT image greyscale component of an image is manipulated via the CT numbers. The CT numbers are represented in terms of Hounsfield Units (HU).
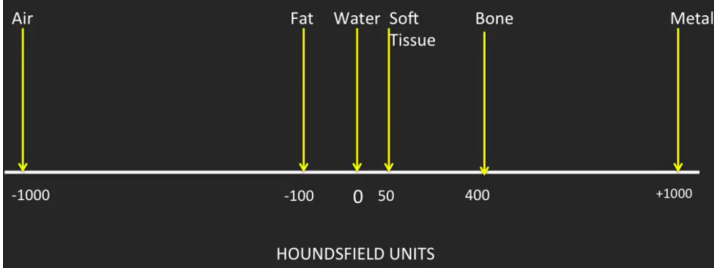

Fig. 2   Hounsfield Units for different mediums.

The CT scans are manipulated using *window width* and *window level*. The window width is the measure of range of different CT numbers. The Window Level (WL) refers to the window centre or midpoint HU value that is represented on the window setting. For our project we use window width of 1800 and window level of 400[12].
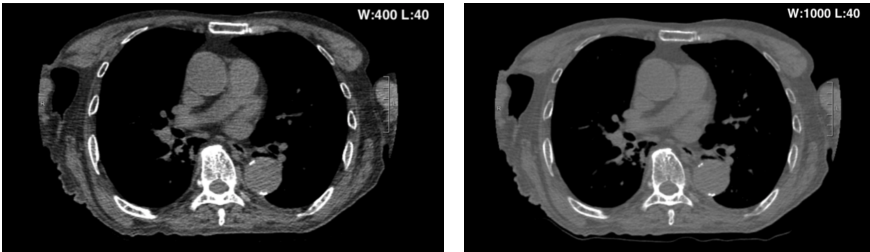

Fig. 3   Impact of window width on the appearnace of a CT scan.

3.2.2 Resizing

Almost all images have size 512x512. We therefore resized the other dimension images to size 512x512.
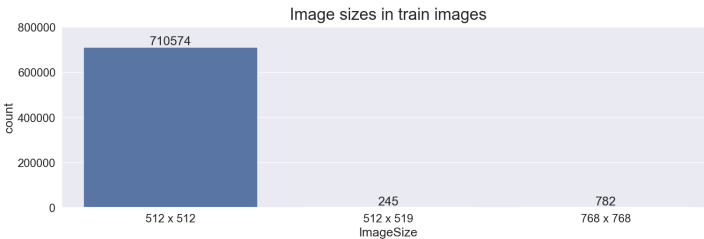

Fig. 4   Image size in train images.

3.3 Detection

We perform fracture detection in the two phases. The first phase is segmentation of all the available study slices of CT radiographs whereas the second phase is fracture detection and localization, which is done by displaying bounding boxes covering fractures. The flow of the training dataset is described in the below figure.
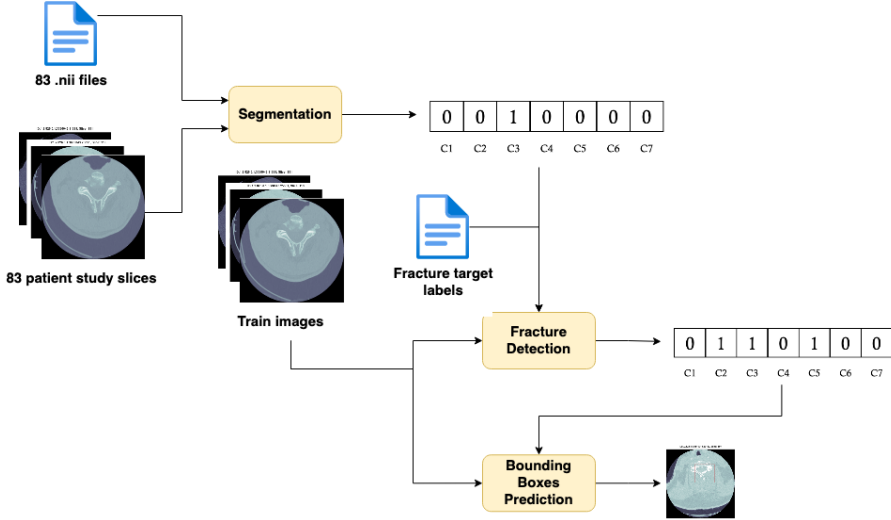

Fig. 5 Data flow in model training.

3.3.1 Segmentation

Segmentation involves classification of CT radiographs into C1 to C7 labels. Using the 83 studies which have segmentation data, we train Random Forest classifier[13] and EfficientNetV2[1]. First we extract vertebrae targets from 83 segmentation files and use them for training. The trained segmentation models take the slices as the input and output the one-hot encoded labels for each slice indicating the vertebrae the slice belongs. Correct vertebrae labels are essential when training the fracture prediction model.

3.3.2 Fracture Detection and Localization

The fracture detection data consists of data indicating which vertebrae has fracture in a particular CT examination (each CT examination contains many slices). Therefore, we store the labels predicted by the segmentation model and merge them with the fracture detection data to obtain the information if a slice in an examination has fracture or not. The model is then trained on this merged data to predict fractures in a particular slice of a CT examination. EfficientNetV2 and YOLOv5[14].

Localization here, refers to highlighting or bounding the fracture with box in the CT radiograph. EfficientNetV2 only produces the fracture detections and not the localization whereas the YOLOv5 model predicts the fractures and localizes the fractures. YOLOv5 model also provides the confidence measure of fracture in the bounding box.

4 Results

4.1 Segmentation

From Table 1, we can see that EfficientNet performs well compared to the Random Forest Classifier. EfficientNetV2 is a new family of convolutional networks that have faster training speed and better parameter efficiency than previous models. To develop these models, a combination of training-aware neural architecture search and scaling was used to jointly optimize training speed and parameter efficiency. The models were searched from the search space enriched with new ops such as Fused-MBConv[15]. Experiments show that EfficientNetV2 models train much faster than state-of-the-art models while being up to 6.8x smaller. With transfer learning, EfficientNetV2 significantly outperforms previous models on ImageNet and CIFAR/Cars/Flowers datasets[1]. By pretraining on the same ImageNet21k, EfficientNetV2 achieves 87.3% top-1 accuracy on ImageNet ILSVRC2012, outperforming the recent ViT by 2.0% accuracy while training 5x-11x faster using the same computing resources.

Vetebrae	Random Forest Classifier	EfficientNetV2
C1	0.8891	0.9621
C2	0.8898	0.9409
C3	0.8962	0.9559
C4	0.8952	0.9605
C5	0.8893	0.9482
C6	0.8713	0.9450
C7	0.8576	0.9465
Overall	0.8841	0.9513

Table 1 Accuracies of the models for the segmentation

The EffNetV2 model is validated using the k-fold ($k = 5$) validation technique. The validation loss and accuracy of the model for k different runs.
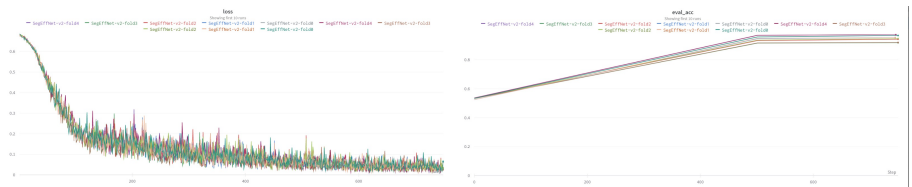

Fig. 6   The validation loss and accuracy of EffNetV2 for differnt runs.

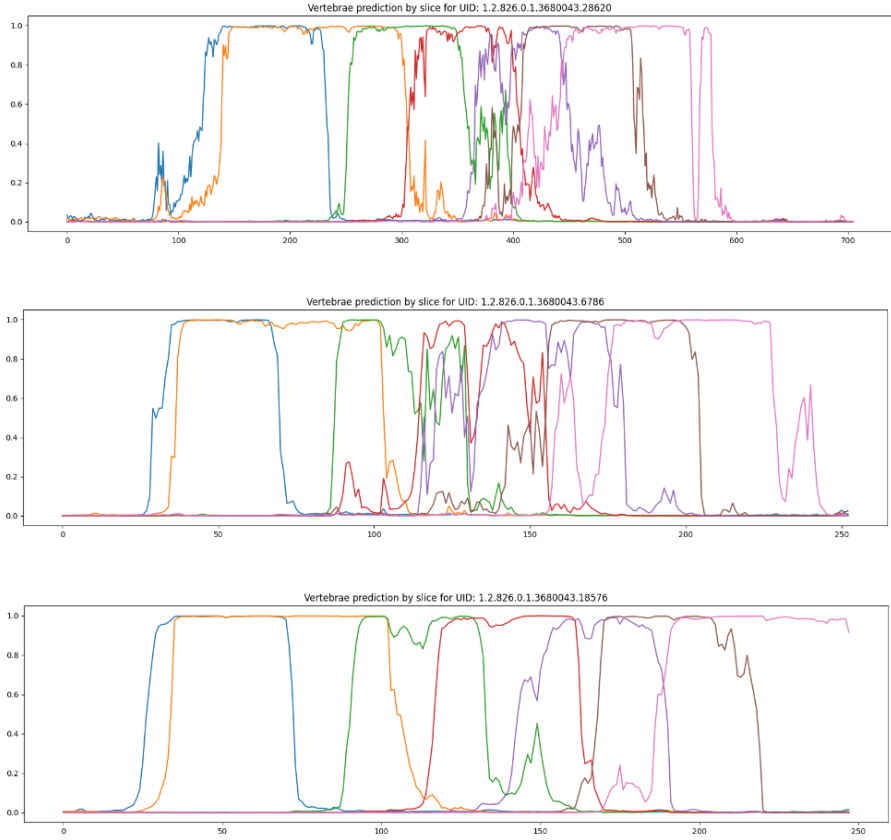Below are the prediction probabilities of 3 study samples for each vertebrae.

Fig. 7 Vertebrae prediction by slice for 3 study samples.

4.2 Fracture Detection

EffNetV2 and YOLOv5 have been trained on the dataset to detect fractures on the CT radiographs. The results of these two models are discussed below

4.2.1 EffNetV2

Need to EffNet results here

4.2.2 YOLOv5

The accuracy acheived by the YOLOv5 model is 94% in fracture detection and localization. Table 2 shows the performance metrics of the model. Similarly, Table 3 shows the average performance of the model.

	Precision	Recall	f1-score	support
Normal	0.73	0.84	0.78	200
Fractured	0.98	0.96	0.97	1444

Table 2   Performance metrics

	Precision	Recall	f1-score	support
Accuracy				1644
Macro Avg	0.85	0.90	0.88	1644
Weighted Avg	0.95	0.94	0.95	1644

Table 3   Average performance metrics

5 Future Works

Future works can implement include fracture highlighting which will create a highlight on top of the predicted bone crack. We intend to use the study to detect cracks in concrete structures using deep learning and image manipulation techniques[16].

On the other hand, current flow of prediction can be improved by creating an ensemble of two models or by create a single model to perform segmentation, fracture detection and localization tasks. Further improvements can be made in segmentation model to impose monotonicity on the predictions i.e., to reduce predictions where the slice is predicted to be in two or more vertebrae.

6 Conclusion

We have demonstrated that deep learning models are able to provide good fracture detection predictions and localization from the CT radiographs. The results have shown the feasibility of the above mentioned models to perform cervical spine fracture detection with 95% in segmentation and 94% in fracture detection and localization. This that the CNNs holds promise at both worklist prioritization and assisting radiologists in cervical spine fracture detection on CT radiographs.

References

[1] Tan, M., Le, Q.V.: Efficientnetv2: Smaller models and faster training. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event. Proceedings of Machine Learning Research, vol. 139, pp. 10096–10106. PMLR, ??? (2021). http://proceedings.mlr.press/v139/tan21a.html

[2] Milby, A.H., Halpern, C.H., Guo, W., Stein, S.C.: Prevalence of cervical spinal injury in trauma. Neurosurg Focus **25**(5), 10 (2008). https://doi.org/10.3171/FOC.2008.25.11.E10

[3] Minja, F.J., Mehta, K.Y., Mian, A.Y.: Current challenges in the use of computed tomography and mr imaging in suspected cervical spine trauma. Neuroimaging Clin N Am **28**(3), 483–493 (2018). https://doi.org/10.1016/j.nic.2018.03.009

[4] Olczak, J., Fahlberg, N., Maki, A., Razavian, A.S., Jilert, A., Stark, A., Sköldenberg, O., Gordon, M.: Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop **88**(6), 581–586 (2017). https://doi.org/10.1080/17453674.2017.1344459

[5] Kim, D.H., MacKinnon, T.: Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol **73**(5), 439–445 (2018). https://doi.org/10.1016/j.crad.2017.11.015

[6] Chung, S.W., Han, S.S., Lee, J.W., Oh, K.-S., Kim, N.R., Yoon, J.P., Kim, J.Y., Moon, S.H., Kwon, J., Lee, H.-J., Noh, Y.-M., Kim, Y.: Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop **89**(4), 468–473 (2018). https://doi.org/10.1080/17453674.2018.1453714

[7] Small, J.E., Osler, P., Paul, A.B., Kunst, M.: Ct cervical spine fracture detection using a convolutional neural network. AJNR Am J Neuroradiol **42**(7), 1341–1347 (2021). https://doi.org/10.3174/ajnr.A7094

[8] Thian, Y.L., Li, Y., Jagmohan, P., Sia, D., Chan, V.E.Y., Tan, R.T.: Convolutional neural networks for automated fracture detection and localization on wrist radiographs. Radiol Artif Intell **1**(1), 180001 (2019). https://doi.org/10.1148/ryai.2019180001

[9] Pranata, Y.D., Wang, K.-C., Wang, J.-C., Idram, I., Lai, J.-Y., Liu, J.-W., Hsieh, I.-H.: Deep learning and surf for automated classification and detection of calcaneus fractures in ct images. Computer Methods and Programs in Biomedicine **171**, 27–37 (2019). https://doi.org/10.1016/j.cmpb.2019.02.006

[10] Mawatari, T., Hayashida, Y., Katsuragawa, S., Yoshimatsu, Y., Hamamura, T., Anai, K., Ueno, M., Yamaga, S., Ueda, I., Terasawa, T., Fujisaki, A., Chihara, C., Miyagi, T., Aoki, T., Korogi, Y.: The effect of deep convolutional neural networks on radiologists' performance in the detection of hip fractures on digital pelvic radiographs. Eur J Radiol **130**, 109188 (2020). https://doi.org/10.1016/j.ejrad.2020.109188

[11] Radiology Society of North America. https://www.rsna.org/

[12] Lee, H., Kim, M., Do, S.: Practical window setting optimization for medical image deep learning. CoRR **abs/1812.00572** (2018) 1812.00572

[13] Buitinck, L., Louppe, G.: Scikit-learn: Machine Learning in Python, Random Forest Classifier. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[14] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., Michael, K., TaoXie, Fang, J., imyhxy, Lorna, Wong, C., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, UnglvKitDe, Sonck, V., tkianai, yxNONG, Skalski, P., Hogan, A., Nair, D., Strobel, M., Jain, M.: ultralytics/yolov5: V7.0 - YOLOv5 SOTA Realtime Instance Segmentation. https://doi.org/10.5281/zenodo.7347926. https://doi.org/10.5281/zenodo.7347926

[15] Gupta, S., Tan, M.: EfficientNet-EdgeTPU: Creating Accelerator-Optimized Neural Networks with AutoML. https://ai.googleblog.com/2019/08/efficientnet-edgetpu-creating.html

[16] Golding, V.P., Gharineiat, Z., Munawar, H.S., Ullah, F.: Crack detection in concrete structures using deep learning. Sustainability **14**(13) (2022). https://doi.org/10.3390/su14138117