Deep Learning for Computer Vision
Final Project Presentation
Team: Conbaconcon

Members:賴繹文R09921128 洪崗峻R09922150 袁肇謙R10921061 丁維R09921064 葉彥廷R10942153

1. Introduction

Given the computed tomography (CT)of skull of different patients, we aim to predict whether fractures exist and where they are. Besides, we'll be tackling three potential challenges: dealing with grey-scaled images/different number of CT for each patient and predicting accurate position of fractures.

2. Method

There are two goals for this competition: Case-level accuracy and Centroid-level F1 score. We deploy a single *AlexNet* for feature extraction and a *transformer* for both case labels and the position of fracture (as fig. 1 shows). To be clear, we take all CT images from a patient and encode them into *embeddings* before feeding into the *transformer*. Also, to utilize the spatial information among slices, *position embeddings* are added to the features.

After that, the *encoder* part of *transformer* (and a *MLP*) will produce case labels within [-1, 0, 1], and the *decoder* part of it (and a *MLP*) will give all the predictions in the format of (index, x, y). For more details, we take *Cross Entropy Loss* for the label classification task. Specially, we design a *Consistency Loss* (fig. 2)to constrain the output patterns of the encoder, which means the labels of slices of a patient should be consistent(0 or 1). After receiving the mutual information among slices from the encoder, the decoder can thus yield tuple of index of slice, x axis and y axis iteratively. MSE between ground truth & prediction is taken as loss.

3. Experiment Results

Accuracy	0.87
F1-Score	0.67

4. Conclusion

There are few works about object detection for predicting dots on a grayscale image. We propose a novel, one stage, transformer-based method to solve this challenge, which not only shows the power of the *Attentions*, but also provides some aspect about this prosperous domain.

Fig. 1: Model Architecture

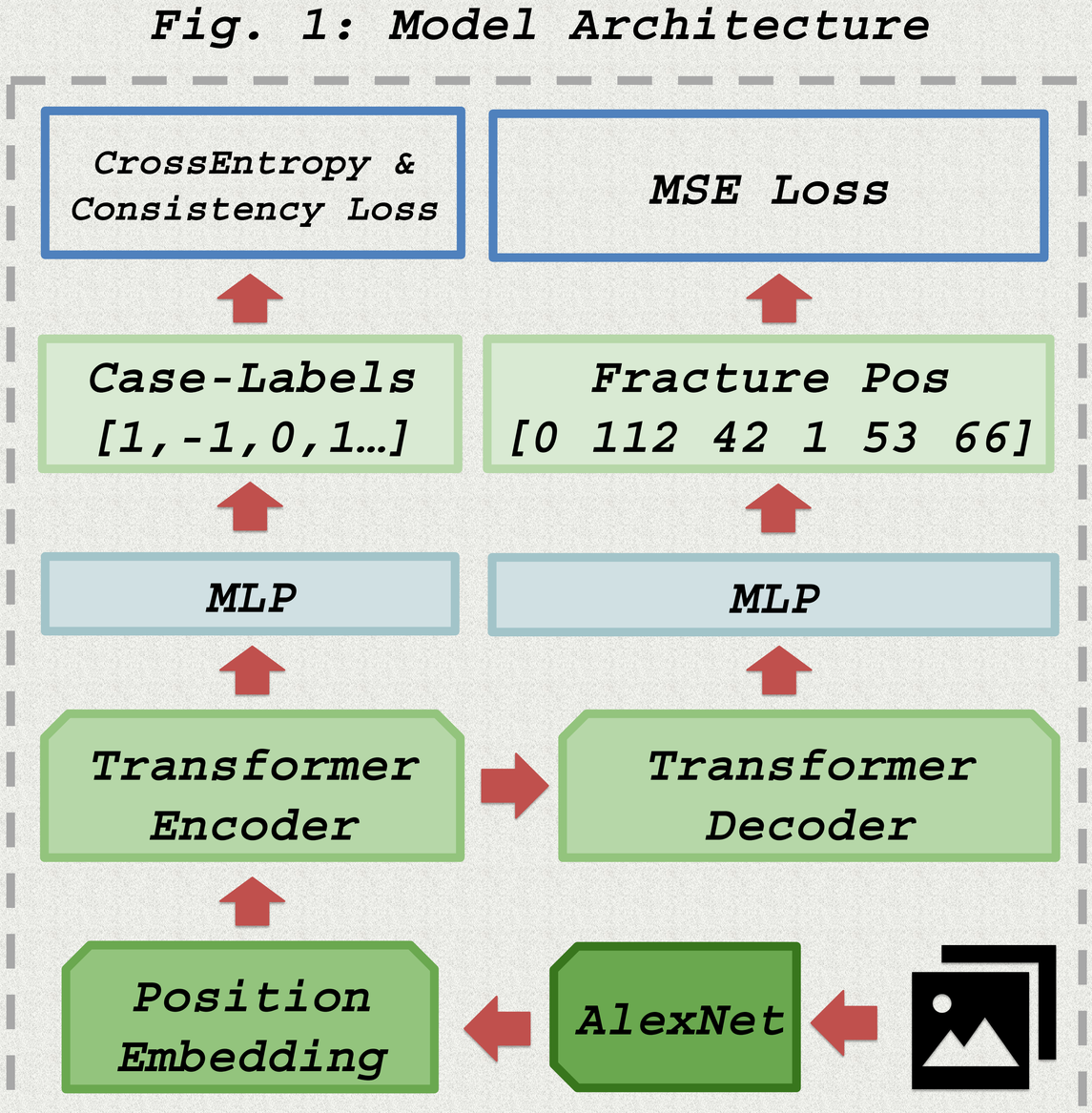

Fig. 2: Consistency Loss