Technical report
Regression CNN in fetal head circumference estimation from ultrasound images and its explainability

Jing Zhang

August 2020

1 Introduction

Fetal head circumference (HC) is one of key biometrics during pregnancy to check whether a fetus growing normally. In clinical diagnostic process, sonographer takes the ultrasound image of fetal head, then draws the outline of fetal head similar to an ellipse, so the HC is the length of the ellipse.

With the computer vision and deep learning techniques rising, automatic measurement of HC is becoming a trending. Conventional methods consists of two steps, one is getting the segmentation results of fetal heads through supervised learning architectures. Another one is fitting the segmentation results into standard ellipses in order to compute HC.

In our work, we propose regression CNN to directly predict HC value without segmenting. The model including CNN and regressional layer two parts which learn the features of fetal head from ultrasound images and predict HC values respectively. We test several models with 3 different regression loss functions, they are two custom regression CNNs and pretrained VGG16 and ResNet50 based regression CNNs, Mean Absolute Error loss, Mean Square Error loss and Huber Loss, respectively.

No matter CNN model or regression CNN model, they are acting like a black box. So we do not know if the model can learn something from images. To get insight of regression models, we investigate the saliency maps of the model by using different explainable methods.

2 Existed problems

2.1 HC Prediction error

As recorded in some literature [1], the HC prediction error based on segmentation and fitting method is around 2 mm (See Table 1, Trimester 2). The

HC prediction error of proposed method is around 4 mm. However, the experiment protocol is different from one to another, for example, the dataset and experiment setting.

2.2 The relationship between HC and trimester

As recorded in literature [1], they compared the HC results in 3 different trimesters. See Table 1. However, it seems not fair because the amount of training data and test data are different from trimester 1,2,3.

Table 1: Results of quantification system C for training set and test set compared to observer 1

	Trimester 1		Trimester 2		Trimester 3	
	Train (165)	Test (55)	Train (693)	Test (233)	Train (141)	Test (47)
DF(mm)	-0.7±6.4	-0.3±6.1	0.7±2.1	0.8±3.3	1.0±6.1	0.6±5.9
ADF(mm)	3.4±5.5	3.1±5.2	2.3±2.1	**2.4±2.4**	4.6±4.1	4.8±3.4
HD(mm)	2.1± 2.7	1.7 ±2.3	1.7±0.9	1.8±1.3	3.4±2.0	3.1±1.9
DSC(%)	93.2 ±7.5	94.4±5.5	97.6±1.8	97.6±1.4	97.2±1.5	97.3±1.5

2.3 Explainability on regression CNN

The supplementary experiment including comparing AOPC scores of two regression models with 3 loss functions and saliency maps of different analysis methods. See Table 2 and Fig. 1.

Table 2: Performance (AOPC scores) of different analysis methods after perturbation, with two regression models and three loss functions. G: Gradient, SG: SmoothGrad, DCN: DeConvNet, DT: DeepTaylor, GB: GuidedBackprop, I*G: Input*Gradient, IG: IntegratedGradients. Lower is better. Best scores in bold.

Model	G	SG	DCN	DT	GB	I*G	IG	LRP
VGG16_MAE	-7.312	-7.398	-2.869	-7.401	-1.663	**-9.189**	**-9.490**	**-9.175**
VGG_MSE	-7.805	-7.181	-5.356	-9.098	-2.990	**-14.568**		**-14.460**
VGG_HL	-23.389	-21.626	-24.588	-27.779	-18.859	**-29.468**		**-29.268**
ResNet50_MAE	-11.533	-11.841	-9.249	-9.890	-9.717	**-14.748**	-5.603	**-14.577**
ResNet50_MSE	-11.305		-11.187	-19.408		**-32.487**	**-20.495**	**-32.505**
ResNet50_HL	-24.174	-24.269		-22.665	-28.421	**-37.125**	-22.814	**-38.120**

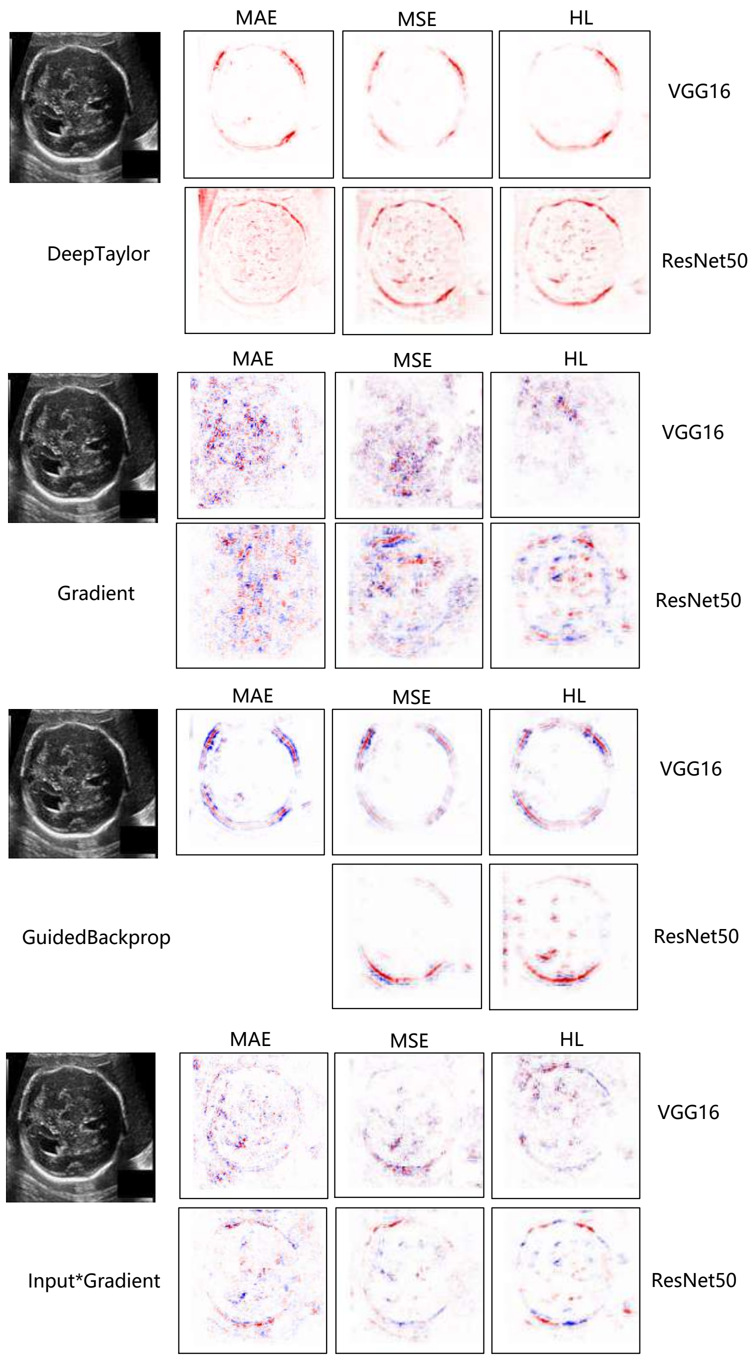

Figure 1: Saliency maps of different explanation methods under 3 different loss functions and regression CNN model VGG16 and ResNet50

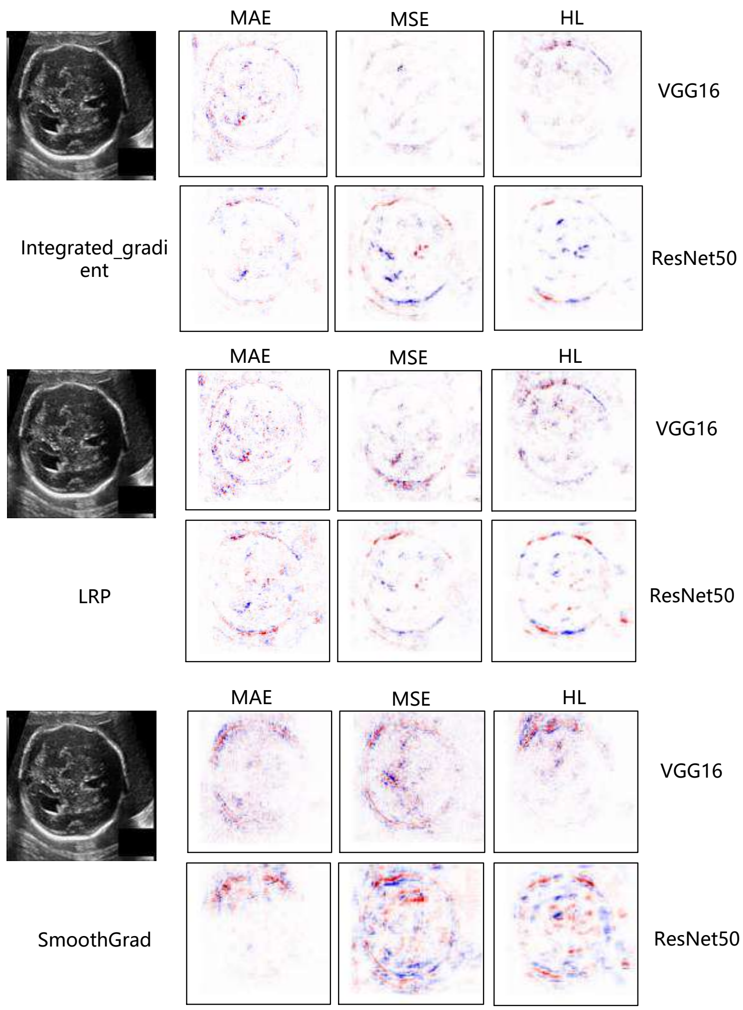

Figure 2: Saliency maps of different explanation methods under 3 different loss functions and regression CNN model VGG16 and ResNet50

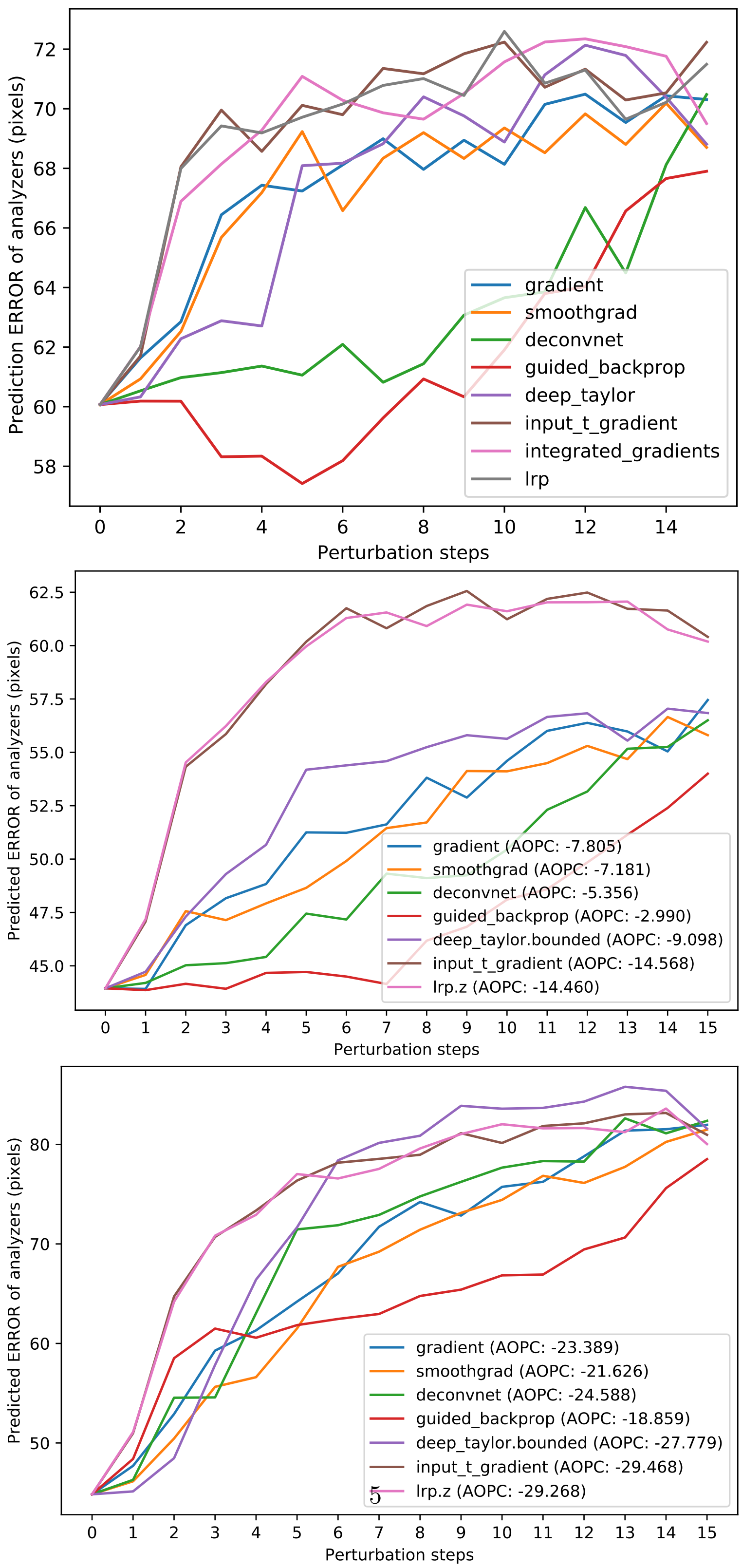

Figure 3: Perturbation steps of different analyzers under Regression VGG16 model with loss function MAE, MSE, HL, repectively.

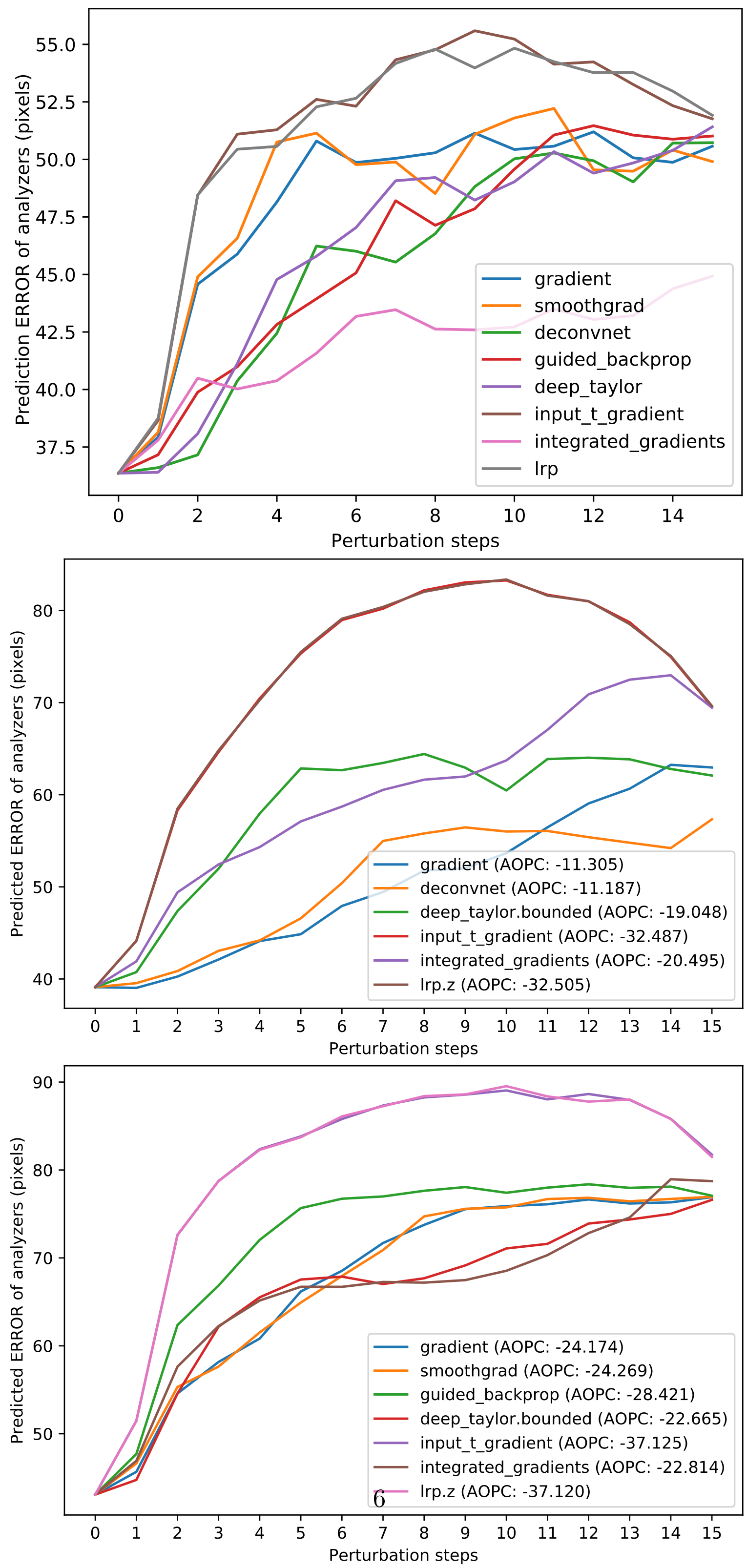

Figure 4: Perturbation steps of different analyzers under Regression ResNet50 model with loss function MAE, MSE, HL, repectively.

3 Summary

For the HC prediction problem, we proposed regression CNN which can directly predict the HC value. The mean absolute error result is 4 mm (average results of three trimesters). And we studied its explainability through several analysis methods in the aspect of data (the image that have good prediction or bad prediction), model (regression VGG16 or regression ResNet50) and regression loss function (MAE, MSE, HL). We explained the networks in the form of saliency maps (visually) and perturbation analysis (quantitatively) to make readers convinced that the regression CNN is able to learn the key features from input data (images) and there are also difference performance among different models, loss functions.

References

[1] Thomas L. A. van den Heuvel, Dagmar de Bruijn, Chris L. de Korte, and Bram van Ginneken. Automated measurement of fetal head circumference using 2D ultrasound images. *PLOS ONE*, 13(8):e0200412, 2018.