BayeRegX: Uncertainty Driven, Attention-based Keypoint Regression for Mensuration Analysis of Spinal X-rays

Uddeshya Upadhyay
Synapsica AI
uddeshya.upa@synapsica.com

Badrinath Singhal
Synapsica AI
badrinath.singhal@synapsica.com

Kuldeep Singh
Synapsica AI
kuldeep@synapsica.com

Meenakshi Singh
Synapsica AI
meenakshi@synapsica.com

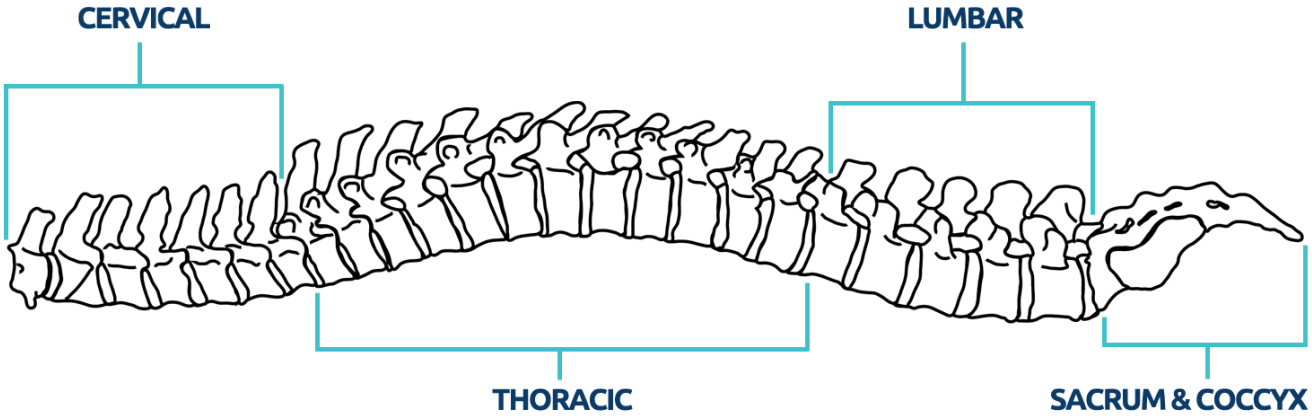


Figure 1: Breakdown of spinal cord in homo-sapiens (image taken from here)

ABSTRACT

Spinal injuries are one of the leading causes of chronic pain for a diverse set of populations. Post spinal trauma, the ligaments of the spine may be damaged causing ligamentous instability which results in abnormal intersegmental motion and Excessive Joint Motion (EJM) that causes acute spine pain and is a significant risk factor for developing chronic pain [8]. Diagnosis of such conditions often requires the acquisition of medical scans such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and X-rays, but these imaging techniques are often unable to clearly show the real source of pain hence unable to determine the correct course of treatment. Different injuries may be graded differently and hence require different treatment protocols. Recently, advanced computerized techniques for measuring X-rays have been used to precisely and accurately determine the grade of spinal injuries [55], such a method has also led to an improved consensus among doctors regarding the grade of injuries. A key component in such computerized methods is the detection of certain keypoints that allows the measurement of various distances and angles which can be used to detect and diagnose different conditions. Such keypoints detection in the scan is currently performed by board-certified radiologists. However, this is a repetitive and laborious task and requires significant time and cost. In this work, we propose a method to automatically detect such keypoints in the scan that can help automate the entire pipeline including the task of grading different spinal injuries and determining the diagnosis. We further propose a method to perform uncertainty quantification in the keypoint prediction using Bayesian deep learning techniques that can indicate how reliable the model predictions are at inference time, allowing human intervention when necessary and avoid fatal consequences. We compare different methods performing key-point detection in X-rays (including the standard techniques without uncertainty estimation and more advanced techniques involving attention and uncertainty estimation) and our experiments show that incorporating attention along with uncertainty estimation leads to state of the art performance. The proposed method is first work which quantifies uncertainty in key point predictions in X-ray.

CHIL '20, April 02–04, 2020, Toronto, Canada

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS CONCEPTS

• **Computing methodologies** → *Computer vision problems.*

KEYWORDS

Keypoint Regression, Bayesian Deep Learning, Uncertainty Estimation, Attention, Spinal X-ray

ACM Reference Format:
Uddeshya Upadhyay, Badrinath Singhal, Kuldeep Singh, and Meenakshi Singh. 2020. BayeRegX: Uncertainty Driven, Attention-based Keypoint Regression for Mensuration Analysis of Spinal X-rays. In *CHIL '20: ACM Conference on Health, Inference and Learning, April 02–04, 2020, Toronto, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Spinal injuries are among the leading causes of chronic pain in a vast diversity of populations. A significant number of patients after spinal injuries may end up with a pain that lasts indefinitely [8, 47]. Often ligaments in the spine are damaged post a spinal trauma causing ligamentous instability that results in abnormal intersegmental motion. Excessive joint motion (EJM) caused by spinal ligament sprains is a cause of acute spine pain and is a significant risk factor for developing chronic pain [8, 39].

To reduce the risk of developing chronic pain post injuries, it is critical that the physician identifies the physical cause of pain in the patient. X-ray, CT, and MRI scans are frequently used to aid the physician in the diagnosis, but often these techniques do not indicate the real cause of pain and serious spinal intersegmental motion problems are routinely missed by board-certified radiologists making the diagnosis more challenging and uncertain [8]. Another factor that makes the diagnosis of spinal injuries even more challenging is the lack of consensus among board-certified radiologists in grading the spinal injuries as the interpretations of the scans are very subjective [8].

In order to overcome these challenges, some recent computerized techniques have been proposed to automate certain parts of the diagnostic procedures for spinal injuries. Many such methods identify the spinal injury conditions not through the physical examination but by measuring X-rays thus providing standardized, precise and objective evaluations. Computerized Radiographic Mensuration Analysis (CRMA) is one such technique that is an accepted procedure in the National Guideline Clearinghouse (NGC) that provides useful clinical information to accurately identify ligament instabilities [55], which in turn determines the treatment plan.

An essential component in such techniques including CRMA is the detection of certain keypoints in the X-rays. These key-points are later used to extract certain lengths and angles which indicate the grade of injury and may help suggest appropriate treatments. Figure 2,3 show the example of the keypoints which are typically required in the analysis of cervical and lumbar X-ray scans.

Currently, the above-mentioned keypoints in spinal X-rays are detected via trained radiologists. Our contributions in this work are manifold. Firstly, we propose a method to automatically detect these keypoints in the scan which can be used for further downstream analysis leading to the final treatment plan. Secondly, we study the performance of the models after incorporating convolutional block attention, which allows our model to focus on important regions in the feature maps. Furthermore, we explore uncertainty quantification in this task of keypoint detection in X-rays using Bayesian deep learning techniques, that not only predict the keypoints but also quantify how confident the model is about a particular prediction. Currently, the widely deployed machine learning models for various applications are often treated as black-box and there is no indication if model predictions are wrong or unreliable, this could lead to an undesirable situation when the model fails. Ability to quantify the uncertainty in the model predictions is of immense value in various applications including medical data-science as it can prevent fatal consequences by allowing human intervention at an appropriate time. In real-world clinical settings, data at inference time could be significantly different from the type of data used for training due to various reasons, for instance, in the case of X-ray scans there could be significant variation due to multiple factors including the type of scanner, position/posture of the patient, the body part being scanned, the physique of the patient, and contrast of the scan, etc. At times, such significant differences in data at inference time can make it harder to detect the keypoints, and having an uncertainty measure could help in introducing human intervention proactively when the model indicates high uncertainty, preventing any fatal consequences. We also propose a scheme to assign the uncertainty to predicted points and perform an empirical study to interpret the uncertainty values provided by the model. Our experiments on a large real-world dataset show that the proposed method yields state of the art performance on the key-point detection in spinal X-rays and also provides meaningful uncertainty estimates for the prediction. The rest of the paper is structured as follows: first, we briefly go over the related work in section 2, then we explain proposed methodologies and describe the dataset used in section 3, then we present results of our experiments in section 4 and finally conclusions and future work in section 5.

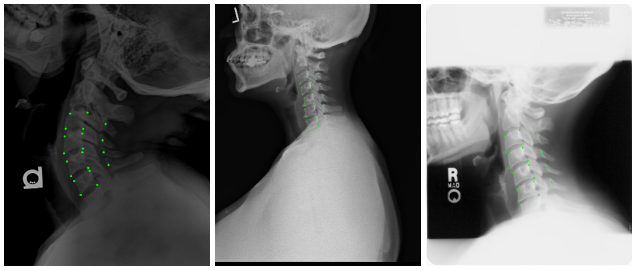


Figure 2: Samples of cervical scans with the point annotations from the dataset

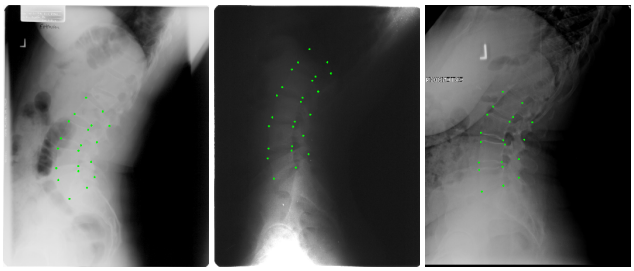


Figure 3: Samples of lumbar scans with the point annotations from the dataset

2 RELATED WORK

Recent times have witnessed a surge of deep learning based techniques to solve a variety of problems in computer vision such as classification [11, 51], localization [18, 73], detection [21, 38, 72], inpainting [68, 69], keypoint detection [1, 30, 52], pose-estimation [35, 37, 56], segmentation [28, 42, 43] and more. Medical image analysis and healthcare is no exception, and has also seen rapid development of such solutions to solve various problems.

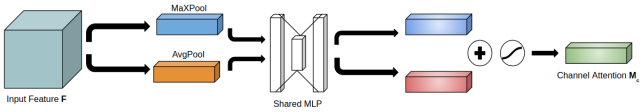
2.1 Deep Learning for Healthcare

Deep learning based methods are now widely used to solve various problems in medical image analysis and more broadly in healthcare. Recent solutions to some standard problems in medical image analysis include deep learning based registration [65, 67, 74] (i.e., process of finding correct alignment between images), deep learning based segmentation [34, 62, 74], shape analysis [6, 22, 29], classification [5, 16, 53, 66], counting [17, 36, 60], super-resolution [23, 40, 58], quality-enhancement [14, 57, 61]. Moreover, recent methods have also explored the application of deep learning in other health informatics domains such as bioinformatics, wearable devices for health monitoring, and more [44]. Some recent solutions also try to reduce the required gadolinium dose in contrast-enhanced brain MRI [20]. Cardiac MRI analysis has also seen significant improvement by the use of deep learning techniques [3, 15, 41]. Other imaging modalities such as CT scans, Ultrasound, X-rays, etc have also benefited immensely by deep learning [9, 13, 33]. In context of spinal image analysis (using different modalities), some of the recent works have focused on segmentation in CT scans [48, 54], automatic opportunistic osteoporosis screening in CT [75], Bone strength prediction [59], detection of sclerotic spine metastases [46], spine and pelvis detection in frontal X-ray [2], localization and identification of vertebrae in spine [10]. Some of the older techniques involved classical handcrafted features and various classifiers like Support Vector Machines (SVM) [63]. Recently proposed [31] performs the segmentation of the discs which allows them to process the axial and sagittal scans of a disc on various levels simultaneously using a pair of deep neural networks to grade the cross-sectional region of discs based on the seriousness of stenosis.

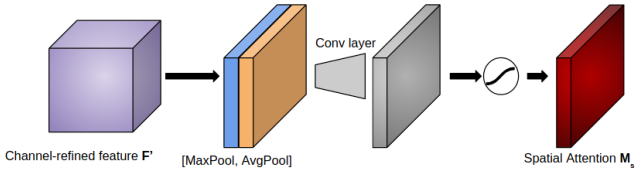
However, none of the previous work has explored automating mensuration analysis in spinal images. This work presents one such study where different methods are tested to perform mensuration analysis in spinal X-ray scans. While we solve a specific problem of detecting key-points in the vertebrae, proposed methods are generic and can be adapted to other use cases trivially.

2.2 Attention Mechanism in Deep Learning

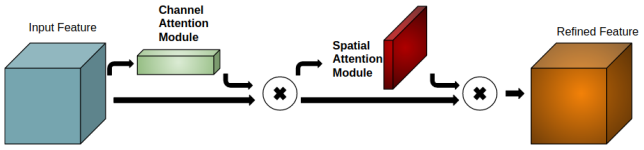
Attention mechanism for deep learning was first introduced in [4, 32] for neural machine translation using sequence-to-sequence (Seq2Seq) models based on Recurrent Neural Networks (RNNs). It was proposed as a method to improve upon one of the limitations of RNNs, that is, the fixed-length context vector is unable to capture the information in longer sequences. The attention mechanism proposed in [4] uses the hidden vectors produced by the encoder at every time step to calculate and assign weights to important parts of the input sequences no matter how far away they are from current time-step, this helps improve the performance of seq2seq



(a) Spectral/Channel attention does pooling over spatial dimensions to derive attention over channels using MLP



(b) Spatial attention pools over channel dimension of the refined feature maps to derive attention over spatial dimensions



(c) Pipeline of CBAM, channel and spatial attention in sequence

Figure 4: Description of various components of Convolutional Block Attention Module (CBAM)

models on longer sequences. Such attention mechanisms are used routinely in various other tasks based on RNNs.

Recently attention mechanisms are proposed for convolutional neural networks as well so that they can be used in image processing pipelines such as salient object detection [7], long-term visual place recognition [12], Spatio-temporal attention mechanism for video captioning [70], attention-based crowd counting network [49]. In this work we use the Convolutional Block Attention Module (CBAM) as proposed in [64] that applies two kinds of attention, spectral and spatial in sequence on a convolutional feature map as shown in figure 4, we describe more on this in section 3.

2.3 Keypoint Regression using Deep Learning

Keypoint regression is an important problem in computer vision where the task is to predict certain keypoints in the image. A variety of problems essentially boils down to keypoint regression, for example, facial landmark detection [71], human pose estimation [56], 6D pose estimation of object [24], hand pose estimation [50].

Typically, convolutional neural networks (CNNs) used for such regression tasks consist of stacked convolutional layers followed by a few fully connected layers with the linear or sigmoid activation function. We take inspiration from such architectures and build our baseline network for the task in a similar manner. The baseline network is then extended to produce state of the art models.

2.4 Bayesian Deep Learning and Uncertainty Estimation

Conventional deep learning systems are good function approximator, that is, they can learn a function that maps an input domain to output domain, where the input-output domain can vary a lot depending upon the tasks such as images-to-label for classification

and image-to-continuous variable for regression problems in computer vision. However, a machine learning model that is deployed in the real world faces a number of challenging tasks including out-of-distribution input, corrupted input, or genuinely harder input where making any prediction is harder, etc. In such cases, the model is bound to give a prediction from the output distribution that it has seen during the training of the model and moreover model will not be able to indicate how confident/certain it is about a prediction.

In contrast, traditional Bayesian machine learning not only provides a prediction but also the corresponding uncertainty estimate through probability distribution over outcomes. However, such techniques have not been widely used in modern machine learning pipelines due to implementation challenges and excessive training times. The work in [19] introduced the method of estimating uncertainties with Deep Neural Networks (DNNs) by training the network with dropouts and using Monte Carlo (MC) samples of predictions using dropout at test time. However, one limitation of this method is that it requires multiple forward passes at the test time, which makes it considerably slower. Also, uncertainty derived from MC samples of predictions only captures the uncertainty in the weights of the network (also known as *epistemic* uncertainty).

Recently, in this work [25], the authors proposed a method to estimate both the inherent uncertainty in the data (known as *Aleatoric* uncertainty) and the *epistemic* uncertainty using DNNs. While the *epistemic* uncertainty is still estimated by MC samples of predictions using dropouts at test time, *Aleatoric* uncertainty is input dependent and learned during the training phase. In big data regime, modeling *Aleatoric* uncertainty is more effective as epistemic uncertainty can be explained away given enough data [25]. We elaborate more on this in section 3.

3 METHODOLOGIES AND DATASET

In this section, we describe the real-world dataset used in this study along with the formal description of the problem that is being solved and the different components used in the solution.

3.1 Dataset

The proprietary dataset used in this study was obtained from a US-based medical device company that partners with spine surgeons to develop motion preservation systems for treating degenerative diseases of the spine. The dataset consisted of $\sim 18K$ spinal X-ray scans out of which $\sim 12K$ are cervical spine and $\sim 6K$ are lumbar spine. The scans were collected from a diverse set of locations in the US using a variety of scanners and patients belong to vast demographics. Board-certified radiologists were used for tagging the dataset by marking the keypoints in various vertebrae as shown in figure 2,3. Both cervical and lumbar spine are significantly different and even within each class, there is a large variation in physical structures of different vertebrae as explained below.

3.1.1 Cervical Spine. This part of the spine is formed by vertebrae between the base of the skull and top of the shoulders. Cervical vertebrae are labeled as C1, C2, and C3-C7 (typically there are 7 cervical vertebrae) as shown in figure 5[1]. The cervical vertebrae protect the spinal cord and work with muscles, tendons, ligaments,

[1]image taken from http://johnhawks.net/explainer/laboratory/types-of-vertebrae.html

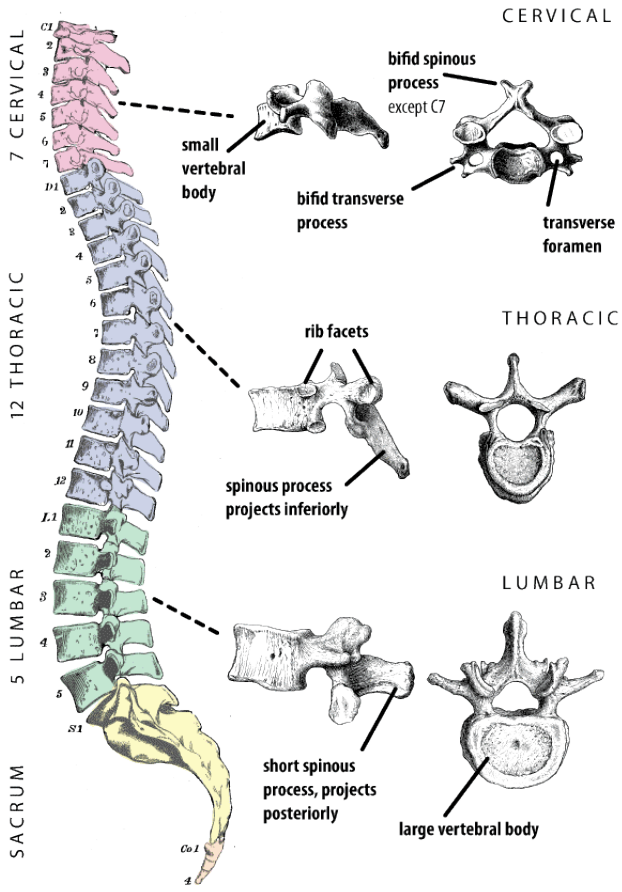


Figure 5: Anatomy of the human spine and vertebrae. There are significant physical differences between different cervical vertebrae and between a cervical and lumbar vertebra.

and joints to provide a combination of support, structure, and flexibility to the neck. The first two vertebrae C1 and C2 are known as *atlas* and *axis* respectively and are highly specialized. They are also known as atypical vertebrae because of significant physical differences compared to other vertebrae. C1 (*atlas*) is the only vertebra without a vertebral body, instead, it is shaped like a ring that connects with the occipital bone above to support the base of the skull. C2, on the other hand, has a large bony protrusion (the *odontoid process*) that points up from the vertebral body and fits into the ring-shaped *atlas* above it. The *atlas* is able to rotate around the *axis*, which leads to a majority of rotational motion in this joint.

3.1.2 Lumbar Spine. This part of the spine is formed by vertebrae found along the body's midline in the lumbar (lower back) region, the lumbar vertebrae make up the region of the spine inferior to the thoracic vertebrae in the thorax and superior to the sacrum and coccyx in the pelvis as shown in figure 5. The lumbar vertebrae labeled as L1-L5, consist of five individual cylindrical bones that form the spine in the lower back. These vertebrae carry all of the upper body's weight while providing flexibility and movement to

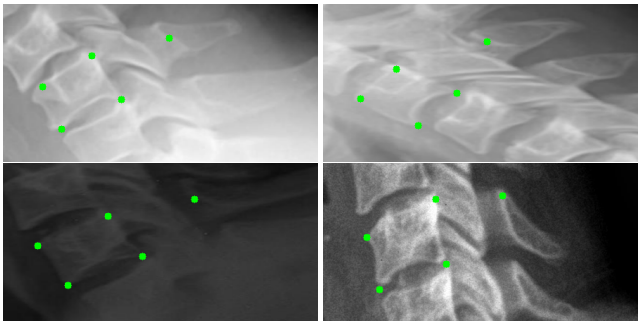


Figure 6: Cervical vertebrae and ground-truth annotations

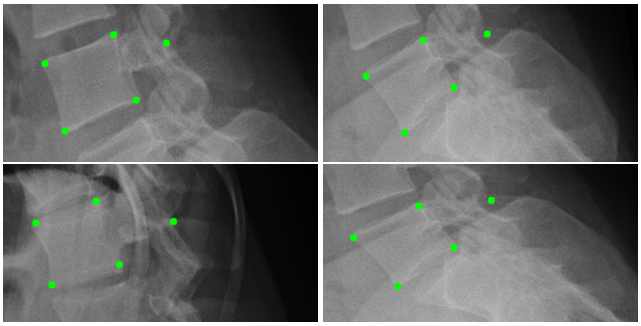


Figure 7: Lumbar vertebrae and ground-truth annotations

the trunk region. They also protect the delicate spinal cord and nerves within their vertebral canal.

In this study, various methods are evaluated on cervical vertebrae C3 to C7 and lumbar vertebrae L1 to L5 as in terms of key-point detection C3-C7 and L1-L5 are very similar with five key-points, four of which are at the corners of the vertebral body and one marking a point on the *spinous process*. Figure 6,7 shows the crops from original cervical and lumbar x-ray scans.

We train different models for predicting keypoints for cervical and lumbar vertebrae. While the original X-ray scans are typical of high resolution, with the entire cervical/lumbar region present, our key-point regression model takes the crop section containing a single vertebra that is obtained using a pre-trained YOLOv3 [45]. The bounding boxes for training YOLOv3 were obtained by creating a box around the vertebra using the points available in the ground-truth. In total, $\sim$ $130K$ number of C3-C7 vertebrae and $\sim$ $55K$ number of lumbar vertebrae were obtained out of which $100K$, and $45K$ were used for training. $10K$, and $5K$ for validation and rest for testing. Let the following represent different sets of vertebrae used for training the various models:

- Set of C3-C7 (collectively represented as Cn) vertebrae by $S_n := \{X_{n,1}, X_{n,2}, ...X_{n,m}\}$
- Set of L1-L5 (collectively represented as Ln) vertebrae by $S'_n := \{X'_{n,1}, X'_{n,2}, ...X'_{n,r}\}$

The following annotations are available for each of the samples from above-defined sets

- $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)\}\ \forall\ X_{n,i} \in S_n$ as there are 5 key-points for C3-C7 (Cn) vertebra

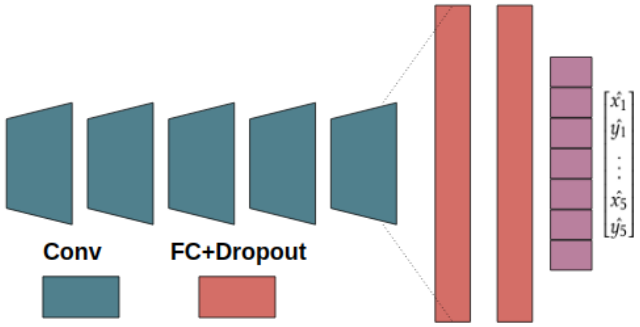


Figure 8: Baseline model to perform keypoint regression based on AlexNet that is a light-weight deep neural network

- $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)\}\ \forall\ X'_{n,i} \in S_n$ as there are 5 key-points for L1-L5 (Ln) vertebra

Where (x_i, y_i) are fractional coordinates (i.e., coordinates in terms of fraction of width and height of the image) of keypoints.

3.2 Architectures and Loss Functions

The performance of multiple models was evaluated and compared for this study as shown in section 4. All of our models perform the point regression task, where the input is an image of a vertebra from X-ray and the output is the X and Y coordinates of the key-point in a pre-specified order. Our models are broadly classified in three different categories: (i) *Baseline*, (ii) *Baseline + CBAM*, and (iii) *Baseline + BayeCBAM*. Architecture and the loss function used to train each model for C3-C7 (Cn) keypoint regression is described in the following, It is similar for L1-L5 (Ln) keypoint regression.

3.2.1 ***Baseline.*** The baseline model was inspired by the light-weight alexnet [27], it consists of 5 convolutional layers followed by fully-connected and dropout layers as shown in figure 8. The number of output nodes depends on the number of keypoints that need to be detected, in the case of C3-C7 and L1-L5 there are 10 output nodes. The coordinates of a point are predicted in a particular order. Figure 8 shows the baseline model used for C3-C7 key-point regression (a similar model is used for L1-L5 as well).

Let the baseline model be represented by $f_B(\cdot; \theta_B)$, and the input to the model be $X \in S_n$. Let the output of the model be $\hat{Y}$ given by,

$$\hat{Y} = [\hat{x_1}, \hat{y_1}, \hat{x_2}, \hat{y_2}, \hat{x_3}, \hat{y_3}, \hat{x_4}, \hat{y_4}, \hat{x_5}, \hat{y_5}] = f_B(X, \theta) \quad (1)$$

Ground-truth annotations represented by,

$$Y = [x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x_5, y_5] \quad (2)$$

is used to compute the loss function to train the network. In our experiment we used Mean Square Error (MSE) loss function to train the baseline network, i.e., $\mathbb{L}_{Base} = MSE(\hat{Y}, Y)$.

3.2.2 ***Baseline + CBAM.*** Recently proposed Convolutional Block Attention Module (CBAM) [64] puts spatial and spectral attention on intermediate convolutional feature maps. Using such an attention based scheme has led to improvement in performance of various deep learning systems. We incorporate the same attention module in our baseline network to get a new model represented by $f_{CBAM}(\cdot; \theta_{CBAM})$ as shown in figure 9. Let the intermediate feature

map after the j^{th} convolutional layer be represented by $F^{C_j \times H_j \times W_j}$, CBAM sequentially infers spectral and spatial attention maps as follows,

- 1D channel (spectral) attention map $M_{P_j}^{C_j \times 1 \times 1}$ given by,

$$M_{P_j}^{C_j \times 1 \times 1} = \sigma(MLP(AvgPool_Q(F))) + \sigma(MLP(MaxPool_Q(F))) \quad (3)$$

- 2D spatial attention map $M_{Q_j}^{1 \times H_j \times W_j}$ given by,

$$M_{Q_j}^{1 \times H_j \times W_j} = \sigma(f^{7\times7}([AvgPool_P(F); MaxPool_P(F)])) \quad (4)$$

Where the $AvgPool_Q, MaxPool_Q$ are pooling operation along the spatial dimension and $AvgPool_P, MaxPool_P$ are pooling operation along the channel dimension. MLP refers to the multi-layer perceptron, and $f^{7\times7}(\cdot)$ represents convolutional operation with 7×7 kernel. Entire operation is summarized in figure 4. The final refined feature map $F_r^{C_j \times H_j \times W_j}$ (after the application of attention maps) is obtained by,

$$F_i^{C_j \times H_j \times W_j} = M_{P_j} \otimes F \quad (5)$$

$$F_r^{C_j \times H_j \times W_j} = M_{Q_j} \otimes F_i \quad (6)$$

Where the F_i is the intermediate feature map and $\otimes$ is the element-wise multiplication. Similar to baseline model, the output of the model $\hat{Y}$ corresponding to input X is given by,

$$\hat{Y} = [\hat{x_1}, \hat{y_1}, \hat{x_2}, \hat{y_2}, \hat{x_3}, \hat{y_3}, \hat{x_4}, \hat{y_4}, \hat{x_5}, \hat{y_5}] = f_{CBAM}(X, \theta_{CBAM}) \quad (7)$$

Again, ground-truth annotations Y are used to compute the MSE loss function to train the network. Hence, loss function to train the *Baseline + CBAM* network is $\mathbb{L}_{CBAM} = MSE(\hat{Y}, Y)$.

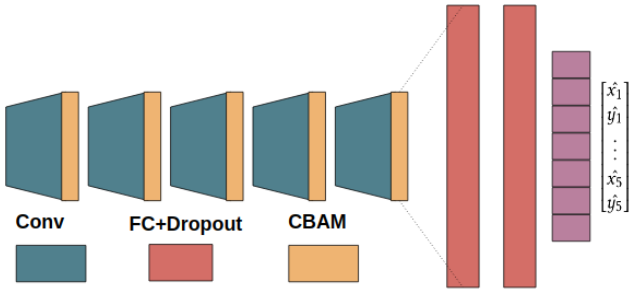


Figure 9: Baseline+CBAM model derived from the Baseline model by introducing convolutional block attention module

3.2.3 ***Baseline + BayeCBAM.*** In order to model *Aleatoric* uncertainty, we modify the model $f_{CBAM}(\cdot; \theta_{CBAM})$ to produce the variance in the prediction (uncertainty) along with the mean values of prediction. The variance is learned in an unsupervised manner as explained below. Let the new model be represented as $f_{Baye}(\cdot; \theta_{Baye})$.

To learn the *Aleatoric* uncertainty (which is input dependent), one must model the variance in the residual obtained by the output of the model and the ground-truth in the following manner. Let the neural network $f'(\cdot; \theta')$ map the set of input $\mathbb{X} = \{X_1, X_2, ...X_m\}$ to the set of output $\mathbb{Y} = \{Y_1, Y_2, ...Y_m\}$,

$$Y_i = f'(X_i; \theta') + \epsilon_i \quad (8)$$

$$\epsilon_i \sim N(0, \sigma_i^2) \quad (9)$$

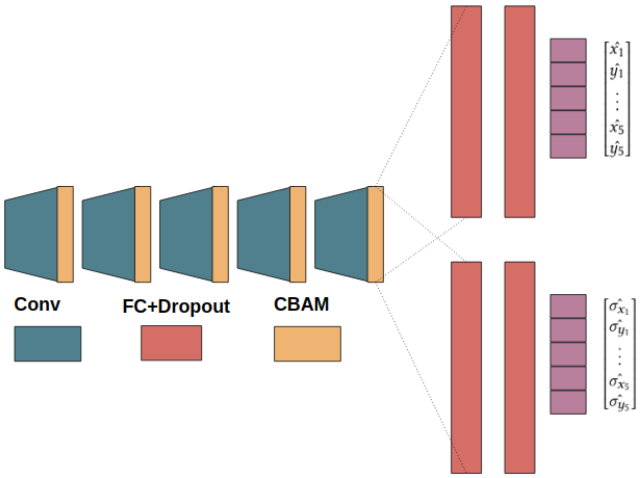


Figure 10: Baseline+BayeCBAM model derived from Baseline+CBAM model by splitting the regression head into two, one for predicting the mean and other for variance

Where ϵ_i is the residual between the prediction and the ground-truth $\forall i \in \{1, 2, ...m\}$ and is assumed to follow a Gaussian distribution with a variance of σ_i^2. Since we are trying to capture the variance inherent in the input, we assume that ϵ_is are independent but the corresponding variance σ_i^2s are unknown and may not be same for all the values of i. The goal is to find the optimal parameter (θ'^*) such that the model maps the input data to output data as closely as possible. One way to do this is to maximize the likelihood function (i.e. maximizing the probability of the data given the parameters, also known as Maximum Likelihood Estimation (MLE)). In practice, maximizing the likelihood is the same as maximizing the log-likelihood, but maximizing the latter makes the optimization simpler. The log-likelihood function is given by,

$$\log(p(\mathbb{Y}|\mathbb{X}, \theta') = \sum_{i=1}^{i=m} \log(p(Y_i|X_i, \theta', \sigma_i)) \quad (10)$$

$$= \sum_{i=1}^{i=m} \log(N(Y_i; X_i, \theta', \sigma_i)) \quad (11)$$

$$= \sum_{i=1}^{i=m} \log(\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(-\frac{||Y_i - \hat{Y}_i)||^2}{2\sigma_i^2})) \quad (12)$$

$$= -\frac{m}{2} \log(2\pi\sigma_i^2) - \sum_{i=1}^{i=m} \frac{||Y_i - \hat{Y}_i||^2}{2\sigma_i^2} \quad (13)$$

Where $\hat{Y}_i = f'(X_i; \theta')$. Equation 13 implies that in this case, maximizing likelihood to get optimal parameter θ'^* is same as,

$$\theta'^* = \arg\min_{\theta'} \frac{1}{m} \sum_{i=1}^{i=m} (\frac{||Y_i - \hat{Y}_i||^2}{2\sigma_i^2} + \frac{\log \sigma_i^2}{2}) \quad (14)$$

To train the network by minimizing the loss function shown in equation 14, the network must also produce an estimate of σ_i (let the estimate be denoted by $\hat{\sigma_i}$) corresponding to output $\hat{Y}_i$. This is achieved by splitting the regression head of the network into two, one for $\hat{Y}_i$ and other for $\hat{\sigma_i}$. Such an architecture that is used in this

study is shown in figure 10, it is derived from *Baseline + CBAM*, where the regression head is split into two, let us represent this by $f_{Baye}(\cdot;\theta_{Baye})$. Therefore the output of the network for an input $X_l \in S_n$ is given by,

$$\begin{bmatrix} \hat{x_1} \\ \hat{y_1} \\ \vdots \\ \hat{x_5} \\ \hat{y_5} \end{bmatrix}, \begin{bmatrix} \hat{\sigma_{x_1}} \\ \hat{\sigma_{y_1}} \\ \vdots \\ \hat{\sigma_{x_5}} \\ \hat{\sigma_{y_5}} \end{bmatrix} = f_{Baye}(X_i;\theta_{Baye}) \tag{15}$$

And the loss function ($\mathbb{L}_{Baye}$) used to train the network is,

$$\mathbb{L}_{Baye} = \frac{1}{B}\sum_{i=1}^{i=B}\frac{1}{10}\sum_{j=1}^{j=5}\frac{||x_j - \hat{x_j}||^2}{2\sigma_{x_j}^2} + \frac{||y_j - \hat{y_j}||^2}{2\sigma_{y_j}^2} + \log\sigma_{x_j}^2 + \log\sigma_{y_j}^2 \tag{16}$$

Where B is the mini-batch size used to train the network. For each key-point, (x, y) coordinates are predicted along with the uncertainty in those predictions given by (σ_x, σ_y) (it has been found empirically that making the network predict $\log(\sigma^2)$ makes the learning smoother [25]), these uncertainty predictions are used to associate a scalar, denoting the uncertainty value to a point as explained in the following.

3.3 Training Procedure

All the models defined above are trained using adam [26] optimizer where the values of (β_1, β_2) are set to $(0.9, 0.999)$. The batch-size for the experiments was set to 96 and initial learning rate was set to $1e^{-4}$. Cosine annealing was used to schedule the decay of the learning rate over epochs and all the models were trained till convergence. At training time we also perform augmentations to the images such as flipping the image left-to-right, changing the contrast by applying CLAHE and adding random padding while generating the crops of vertebrae. Models were trained using images of size 224×448 and on GeForce RTX-2080 Ti 11GB graphics card by Nvidia. Since the training dataset for lumbar vertebrae was significantly smaller than that of cervical vertebrae, we first trained our model for cervical vertebrae and then used it to initialize the second (identical) model for lumbar vertebrae, this led to state of the art performance for both cervical and lumbar keypoint regression.

3.4 Associating Uncertainty to Points

for every keypoint *Baseline + BayeCBAM* predicts the coordinates (x, y) and the corresponding uncertainty values (σ_x, σ_y), in order to associate a scalar uncertainty value (σ_p) to a point p, we define

$$\sigma_p = \sqrt{\sigma_{x_p}^2 + \sigma_{y_p}^2} \tag{17}$$

Such a scalar value of σ_p can be used to visualize and interpret the uncertainty in the predicted point. In this work we propose one method to evaluate and visualize the uncertainty value. Our scheme is described in the following,

- if $\sigma_p < \tau_1$ then certain
- if $\tau_1 \leq \sigma_p \leq \tau_2$ then moderately uncertain
- if $\sigma_p \geq \tau_2$ then highly uncertain

Here (τ_1, τ_2) are application dependent threshold values which are chosen in a systematic manner as explained in section 4. We visualize the predicted points and the corresponding uncertainty by color-coding the points, where colors are decided on the basis of the class they belong: certain (violet), moderately uncertain (yellow) or highly uncertain (red). Figure 11 summarizes the proposed scheme.

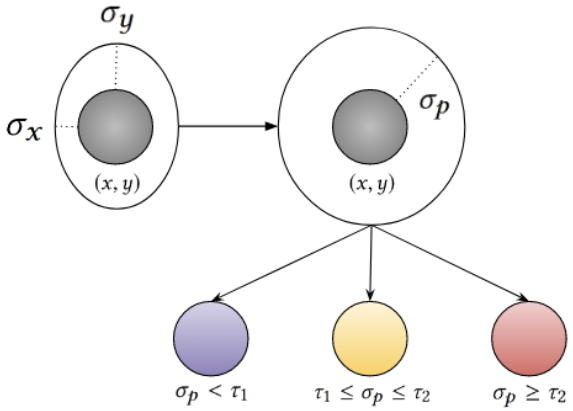


Figure 11: Colouring scheme for visualizing the uncertainty of the points. The scheme depends on two data-dependent parameters (τ_1, τ_2), tuned to limit the error in the predictions, therefore increasing the reliability of the measure

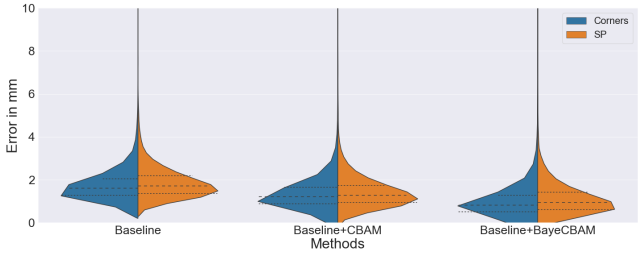


Figure 12: Violin plot comparing performance of all the methods on test dataset

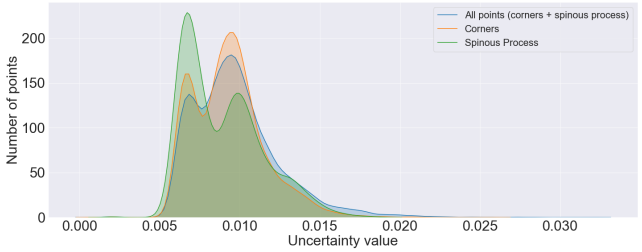
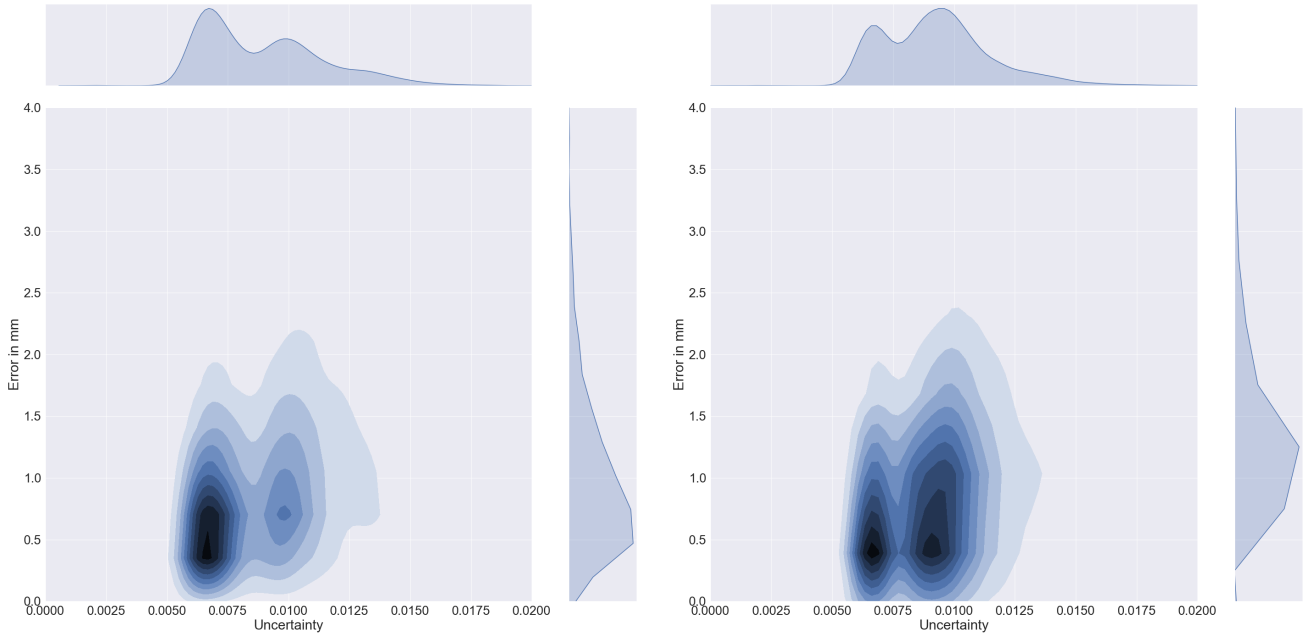


Figure 13: Density function of uncertainty values for the keypoints corresponding to *corners* and *spinous process*

4 RESULTS AND DISCUSSIONS

To measure the performance of our networks, we measure the euclidean distance between the predicted point and the ground truth annotations in millimeters (i.e., error in mm) (using the resolution information present in the scans). Performance of various models are compared (*Baseline*, *Baseline + CBAM*, and *Baseline +*



(a) For spinous processs (SP) keypoints **(b) For corner keypoints**

Figure 14: Bi-variate distribution of (Error, Uncertainty) for cervical test dataset visualized using kernel density estimation

BayeCBAM). Figure 12 shows the violin plot on the test set. A few observations that can be made from the plot are (i) the performance of the *Baseline* model improves when CBAM is added (*i.e., Baseline + CBAM*) and performance is further improved by modifying the framework to estimate uncertainty (i.e., *Baseline + BayeCBAM*). (ii) For a given model, learning to predict the *corner* keypoints in the vertebra crop accurately is relatively easier (less error) compared to the keypoint representing *spinous process* (SP) (relatively higher error). This is aligned with the intuition as the *spinous process* in the images has the highest structural variation and corruptions, making it harder to detect.

Experiments show that for *Baseline* model 89% of scans have average euclidean error less than 2*mm*, for *Baseline + CBAM* 91.2% of scans have average euclidean error less than 2*mm*, and similarly for *Baseline + BayeCBAM* 94.7% of scans have average euclidean error less than 2*mm*, again indicating that *Baseline + BayeCBAM* is superior to other methods. In addition to performing better than other models, *Baseline + BayeCBAM* also provides and interpretable measure of uncertainty indicating how reliable a particular prediction as discussed in section 3. Uncertainty can be high due to various reasons such as keypoint not being clearly visible, low signal to noise ratio, improper contrast, anomalous externalities such as medical devices, occluded keypoint, etc.

It is of paramount importance that the prediction marked as certain by the model is not very far away from the original location of keypoint, therefore we study the correlation between the error in predictions and the uncertainty values. Figure 13 shows the density function (using kernel density estimation (KDE)) of the uncertainty values corresponding to *corner* points and SP points in

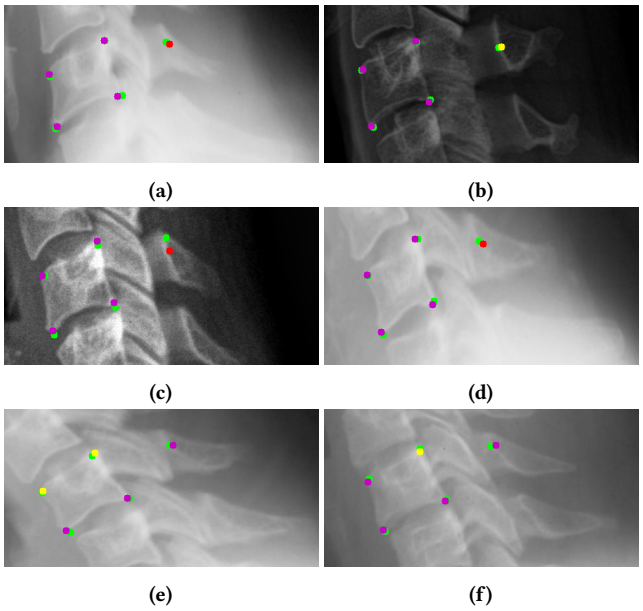


Figure 15: Sample outputs of Baseline+BayeCBAM on cervical vertebrae, where the model is mostly certain (i.e., many keypoints are marked violet)

cervical scans. This helps us understand what range of uncertainty values could be marked as certain/moderately uncertain/highly uncertain. In order to implement the scheme associating uncertainty to individual predicted points explained in section 3, one must systematically tune the data dependent parameters (τ_1, τ_2). Intuition

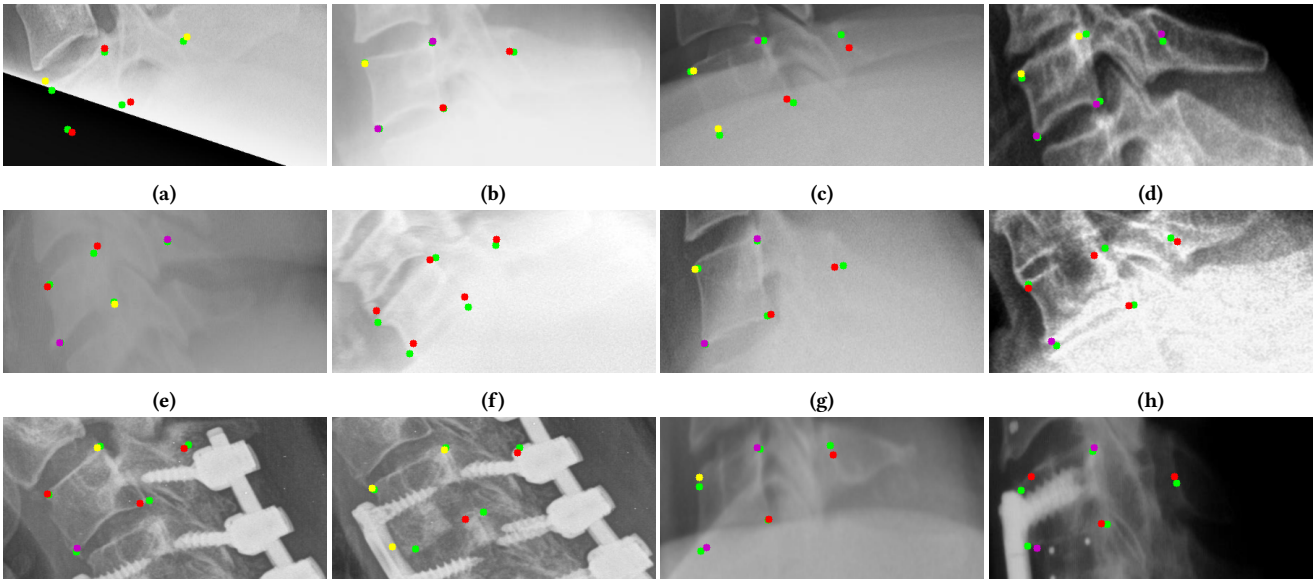


Figure 16: Sample outputs from Baseline+BayeCBAM on cervical vertebrae where the uncertainty is high

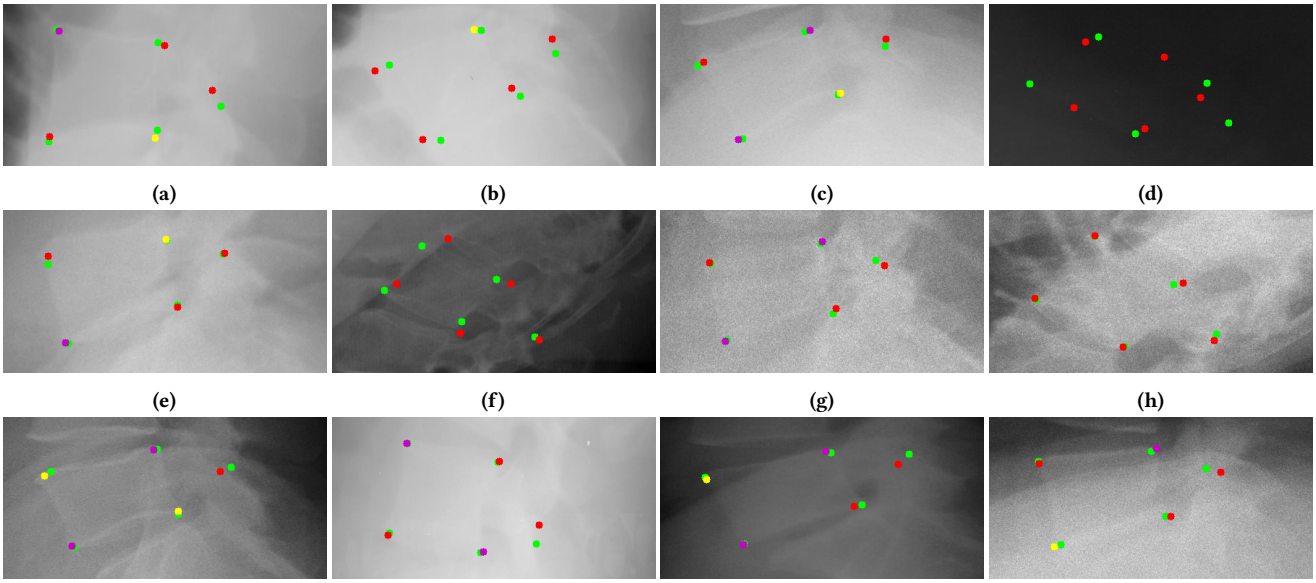


Figure 17: Sample outputs from Baseline+BayeCBAM on lumbar vertebrae where the uncertainty is high

behind picking the values of these parameters can be summarized as follows: (i) τ_1 should be such that, for any given point prediction (p) if $\sigma_p < \tau_1$ then the error in the prediction should not be more than 1.5mm (from figure 14 significant number of scans are included in this region). (ii) τ_2 should be such that for any given point prediction (p) if $\tau_1 \leq \sigma_p < \tau_2$ then the error in the prediction should not be more than 2mm. Figure 14 visualizes the bi-variate distribution of the *Error in mm* and *Uncertainty values* for both *corner* points and *spinous process* using KDE. Using these plots, we set $(\tau_1, \tau_2) = (0.0087, 0.0097)$.

4.1 Qualitative Analysis of Proposed Point Uncertainty Scheme

Figure 15 shows some of the output from the model, where the model is most certain about the predictions for cervical vertebrae.

Points are divided into three categories *certain, moderately uncertain, highly uncertain* marked in violet, yellow and red respectively. Ground-truth annotations are shown in green. Notice that the points marked as certain (i.e., in violet) do not deviate much from the ground-truth annotations and in some of the cases may also be better than noisy ground-truth annotations (i.e., annotations with human error), as shown in figure 15-(a,c).

Figure 16,17 shows the output of the model where the model is most uncertain about the prediction, on the cervical and lumbar dataset respectively. Uncertainty can be high due to various reasons, for instance in 16-a the spinous process is not clearly visible, the image has improper contrast and one of the corner keypoints is not present in the image. Notice that the model was able to roughly estimate the position of the missing keypoint using the position and orientation of other predicted keypoints, but the point is marked highly uncertain, as desired.

Figures 16-(b,c,f,k) and 17-(b,c,j,l) show examples where the quality of the scan is degraded in certain regions close to keypoints making it harder to detect some of the keypoints, which get marked as *moderately* and *highly* uncertain. Notice that if certain keypoints are not affected by the degraded quality of scan in other regions then the point is predicated correctly with high certainty.

Figure 16-(i,j,l) show examples where externalities (medical devices) are present in the image, such externalities are rare phenomena and are not densely covered in the dataset. As a result, model is unable to learn the task properly in the presence of such factors, this is reflected in the predicted output as they are significantly distorted with respect to ground-truth configuration. However, the model is able to indicate that the predicted points are not reliable by marking them with high uncertainty.

5 CONCLUSIONS

In this paper, we present a novel framework to perform keypoint regression for mensuration analysis of spinal X-ray scans. The proposed method can allow the automatic diagnosis of spinal injuries which are among the leading causes of chronic pain. Unlike many machine learning systems that are widely deployed in real-world, our model is designed to predict the uncertainty values along with the prediction, that indicates how reliable are the model predictions, which is of immense value in medical analysis as it can allow human intervention appropriately preventing any fatal consequences. We conduct empirical studies using large clinical dataset and show that our model achieves state of the art performance and also provides a reliable uncertainty estimate.

REFERENCES

[1] Hani Altwaijry, Andreas Veit, Serge J Belongie, and Cornell Tech. 2016. Learning to Detect and Match Keypoints with Deep Architectures.. In *BMVC*.

[2] Benjamin Aubert, Carlos Vazquez, Thierry Cresson, Stefan Parent, and Jacques De Guise. 2016. Automatic spine and pelvis detection in frontal X-rays using deep neural networks for patch displacement learning. In *2016 ieee 13th international symposium on biomedical imaging (isbi)*. IEEE, 1426–1429.

[3] MR Avendi, Arash Kheradvar, and Hamid Jafarkhani. 2016. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical image analysis* 30 (2016), 108–119.

[4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[5] Yaniv Bar, Idit Diamant, Lior Wolf, Sivan Lieberman, Eli Konen, and Hayit Greenspan. 2015. Chest pathology detection using deep learning with non-medical training. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 294–297.

[6] Riddhish Bhalodia, Shireen Y Elhabian, Ladislav Kavan, and Ross T Whitaker. 2018. Deepssm: A deep learning framework for statistical shape modeling from raw images. In *International Workshop on Shape in Medical Imaging*. Springer, 244–257.

[7] Ali Borji, Ming-Ming Cheng, Qibin Hou, Huaizu Jiang, and Jia Li. 2014. Salient object detection: A survey. *Computational Visual Media* (2014), 1–34.

[8] Steven Brownstein, Jefferey Cronk, and Joseph Cioffi. 2015. valuation of Spinal Ligamentous Injuries using Computerized X-ray Interpretation. *Orthop Rheumatol Open Access Journal* (2015).

[9] Kenny H Cha, Lubomir Hadjiiski, Ravi K Samala, Heang-Ping Chan, Elaine M Caoili, and Richard H Cohan. 2016. Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. *Medical physics* 43, 4 (2016), 1882–1896.

[10] Hao Chen, Chiyao Shen, Jing Qin, Dong Ni, Lin Shi, Jack CY Cheng, and Pheng-Ann Heng. 2015. Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 515–522.

[11] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. 2014. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing* 7, 6 (2014), 2094–2107.

[12] Zetao Chen, Lingqiao Liu, Inkyu Sa, Zongyuan Ge, and Margarita Chli. 2018. Learning context flexible attention model for long-term visual place recognition. *IEEE Robotics and Automation Letters* 3, 4 (2018), 4015–4022.

[13] Jie-Zhi Cheng, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. 2016. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Scientific reports* 6 (2016), 24454.

[14] Jianning Chi, Yifei Zhang, Xiaosheng Yu, Ying Wang, and Chengdong Wu. 2019. Computed tomography (CT) image quality enhancement via a uniform framework integrating noise estimation and super-resolution networks. *Sensors* 19, 15 (2019), 3348.

[15] Omar Emad, Inas A Yassine, and Ahmed S Fahmy. 2015. Automatic localization of the left ventricle in cardiac MRI images using deep learning. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 683–686.

[16] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (2017), 115.

[17] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al. 2019. U-Net: deep learning for cell counting, detection, and morphometry. *Nature methods* 16, 1 (2019), 67.

[18] Haoqiang Fan and Erjin Zhou. 2016. Approaching human level facial landmark localization by deep learning. *Image and Vision Computing* 47 (2016), 27–35.

[19] Yarin Gal and Zoubin Ghahramani. 2015. Dropout as a Bayesian approximation: Insights and applications. In *Deep Learning Workshop, ICML*, Vol. 1. 2.

[20] Enhao Gong, John M Pauly, Max Wintermark, and Greg Zaharchuk. 2018. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *Journal of Magnetic Resonance Imaging* 48, 2 (2018), 330–340.

[21] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama* 316, 22 (2016), 2402–2410.

[22] Benjamín Gutiérrez-Becker and Christian Wachinger. 2018. Deep multi-structural shape analysis: application to neuroanatomy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 523–531.

[23] Larissa Heinrich, John A Bogovic, and Stephan Saalfeld. 2017. Deep learning for isotropic super-resolution from non-isotropic 3D electron microscopy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 135–143.

[24] Wadim Kehl, Fabian Manhardt, Federico Tombari, Slobodan Ilic, and Nassir Navab. 2017. SSD-6D: Making RGB-based 3D detection and 6D pose estimation great again. In *Proceedings of the IEEE International Conference on Computer Vision*. 1521–1529.

[25] Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in bayesian deep learning for computer vision?. In *Advances in neural information processing systems*. 5574–5584.

[26] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural*

Information Processing Systems 25. 1097–1105.

[28] Shu Liao, Yaozong Gao, Aytekin Oto, and Dinggang Shen. 2013. Representation learning: a unified deep learning framework for automatic prostate MR segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 254–261.

[29] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical image analysis* 42 (2017), 60–88.

[30] Jonathan L Long, Ning Zhang, and Trevor Darrell. 2014. Do convnets learn correspondence?. In *Advances in Neural Information Processing Systems*. 1601–1609.

[31] Jen-Tang Lu, Stefano Pedemonte, Bernardo Bizzo, Sean Doyle, Katherine P Andriole, Mark H Michalski, R Gilberto Gonzalez, and Stuart R Pomerantz. 2018. DeepSPINE: Automated Lumbar Vertebral Segmentation, Disc-level Designation, and Spinal Stenosis Grading Using Deep Learning. *arXiv preprint arXiv:1807.10215* (2018).

[32] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025* (2015).

[33] Fausto Milletari, Seyed-Ahmad Ahmadi, Christine Kroll, Annika Plate, Verena Rozanski, Juliana Maiostre, Johannes Levin, Olaf Dietrich, Birgit Ertl-Wagner, Kai Bötzel, et al. 2017. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Computer Vision and Image Understanding* 164 (2017), 92–102.

[34] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 565–571.

[35] Arsalan Mousavian, Dragomir Anguelov, John Flynn, and Jana Kosecka. 2017. 3d bounding box estimation using deep learning and geometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7074–7082.

[36] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S Palmer, Craig Packer, and Jeff Clune. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences* 115, 25 (2018), E5716–E5725.

[37] Wanli Ouyang, Xiao Chu, and Xiaogang Wang. 2014. Multi-source deep learning for human pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2329–2336.

[38] Wanli Ouyang and Xiaogang Wang. 2013. Joint deep learning for pedestrian detection. In *Proceedings of the IEEE International Conference on Computer Vision*. 2056–2063.

[39] Manohar M Panjabi. 2006. A hypothesis of chronic back pain: ligament subfailure injuries lead to muscle control dysfunction. *European spine journal* 15, 5 (2006), 668–676.

[40] Chi-Hieu Pham, Aurélien Ducournau, Ronan Fablet, and François Rousseau. 2017. Brain MRI super-resolution using deep 3D convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 197–200.

[41] Rudra PK Poudel, Pablo Lamata, and Giovanni Montana. 2016. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In *Reconstruction, segmentation, and analysis of medical images*. Springer, 83–94.

[42] Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen. 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *International conference on medical image computing and computer-assisted intervention*. Springer, 246–253.

[43] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 652–660.

[44] Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. 2016. Deep learning for health informatics. *IEEE journal of biomedical and health informatics* 21, 1 (2016), 4–21.

[45] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[46] Holger R Roth, Jianhua Yao, Le Lu, James Stieger, Joseph E Burns, and Ronald M Summers. 2015. Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications. In *Recent advances in computational methods and clinical applications for spine imaging*. Springer, 3–12.

[47] Jerome Schofferman, Nikolai Bogduk, and Paul Slosar. 2007. Chronic whiplash and whiplash-associated disorders: an evidence-based approach. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 15, 10 (2007), 596–606.

[48] Anjany Sekuboyina, Jan Kukačka, Jan S Kirschke, Bjoern H Menze, and Alexander Valentinitsch. 2017. Attention-driven deep learning for pathological spine segmentation. In *International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging*. Springer, 108–119.

[49] Vishwanath A Sindagi and Vishal M Patel. 2019. Ha-ccn: Hierarchical attention-based crowd counting network. *IEEE Transactions on Image Processing* 29 (2019), 323–335.

[50] Ayan Sinha, Chiho Choi, and Karthik Ramani. 2016. Deephand: Robust hand pose estimation by completing a matrix imputed with deep features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4150–4158.

[51] Richard Socher, Brody Huval, Bharath Bath, Christopher D Manning, and Andrew Y Ng. 2012. Convolutional-recursive deep learning for 3d object classification. In *Advances in neural information processing systems*. 656–664.

[52] Yi Sun, Xiaogang Wang, and Xiaoou Tang. 2013. Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3476–3483.

[53] Yanmin Sun, Andrew KC Wong, and Mohamed S Kamel. 2009. Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence* 23, 04 (2009), 687–719.

[54] Amin Suzani, Abtin Rasoulian, Alexander Seitel, Sidney Fels, Robert N Rohling, and Purang Abolmaesumi. 2015. Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images. In *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, Vol. 9415. International Society for Optics and Photonics, 941514.

[55] Mehul Taylor, John A Hipp, Stanley D Gertzbein, Shankar Gopinath, and Charles A Reitman. 2007. Observer agreement in assessing flexion-extension X-rays of the cervical spine, with and without the use of quantitative measurements of intervertebral motion. *The Spine Journal* 7, 6 (2007), 654–658.

[56] Alexander Toshev and Christian Szegedy. 2014. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1653–1660.

[57] Uddeshya Upadhyay and Suyash P Awate. 2019. A Mixed-Supervision Multilevel GAN Framework for Image Quality Enhancement. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 556–564.

[58] U. Upadhyay and S. P. Awate. 2019. Robust Super-Resolution Gan, with Manifold-Based and Perception Loss. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*.

[59] Tom Van Steenkiste, Joeri Ruyssinck, Olivier Janssens, Baptist Vandersmissen, Florian Vandecasteele, Pieter Devolder, Eric Achten, Sofie Van Hoecke, Dirk Deschrijver, and Tom Dhaene. 2018. Automated assessment of bone age using deep learning and Gaussian process regression. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 674–677.

[60] Mitko Veta, Paul J Van Diest, Mehdi Jiwa, Shaimaa Al-Janabi, and Josien PW Pluim. 2016. Mitosis counting in breast cancer: Object-level interobserver agreement and comparison to an automatic method. *PloS one* 11, 8 (2016), e0161286.

[61] Thang Vu, Cao Van Nguyen, Trung X Pham, Tung M Luu, and Chang D Yoo. 2018. Fast and efficient image quality enhancement via desubpixel convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 0–0.

[62] Guotai Wang, Wenqi Li, Maria A Zuluaga, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, et al. 2018. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE transactions on medical imaging* 37, 7 (2018), 1562–1573.

[63] Tatjana Wiese, Joseph Burns, Jianhua Yao, and Ronald M Summers. 2011. Computer-aided detection of sclerotic bone metastases in the spine using watershed algorithm and support vector machines. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 152–155.

[64] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 3–19.

[65] Guorong Wu, Minjeong Kim, Qian Wang, Brent C Munsell, and Dinggang Shen. 2015. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE Transactions on Biomedical Engineering* 63, 7 (2015), 1505–1516.

[66] Yan Xu, Tao Mo, Qiwei Feng, Peilin Zhong, Maode Lai, I Eric, and Chao Chang. 2014. Deep learning of feature representation with multiple instance learning for medical image analysis. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 1626–1630.

[67] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. 2017. Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage* 158 (2017), 378–396.

[68] Raymond A Yeh, Chen Chen, Teck Yian Lim, Alexander G Schwing, Mark Hasegawa-Johnson, and Minh N Do. 2017. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5485–5493.

[69] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2018. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5505–5514.

[70] Mihai Zanfir, Elisabeta Marinoiu, and Cristian Sminchisescu. 2016. Spatio-temporal attention models for grounded video captioning. In *asian conference on computer vision*. Springer, 104–119.

[71] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. 2014. Facial landmark detection by deep multi-task learning. In *European conference on computer vision*. Springer, 94–108.

[72] Rui Zhao, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. 2015. Saliency detection by multi-context deep learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1265–1274.

[73] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2921–2929.

[74] S Kevin Zhou, Hayit Greenspan, and Dinggang Shen. 2017. *Deep learning for medical image analysis*. Academic Press.

[75] Timothy J Ziemlewicz, Neil Binkley, and Perry J Pickhardt. 2015. Opportunistic osteoporosis screening: addition of quantitative CT bone mineral density evaluation to CT colonography. *Journal of the American College of Radiology* 12, 10 (2015), 1036–1041.