Hindawi
Journal of Electrical and Computer Engineering
Volume 2018, Article ID 2187247, 13 pages
https://doi.org/10.1155/2018/2187247

Research Article

Versatile Framework for Medical Image Processing and Analysis with Application to Automatic Bone Age Assessment

Chen Zhao,[1,2] Jungang Han,[1,2] Yang Jia,[1,2] Lianghui Fan,[1,2] and Fan Gou[1,2]

[1] *School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China*
[2] *Shaanxi Provincial Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China*

Correspondence should be addressed to Chen Zhao; 1603210019@stu.xupt.edu.cn

Received 9 September 2018; Accepted 12 November 2018; Published 31 December 2018

Academic Editor: Sos S. Agaian

Deep learning technique has made a tremendous impact on medical image processing and analysis. Typically, the procedure of medical image processing and analysis via deep learning technique includes image segmentation, image enhancement, and classification or regression. A challenge for supervised deep learning frequently mentioned is the lack of annotated training data. In this paper, we aim to address the problems of training transferred deep neural networks with limited amount of annotated data. We proposed a versatile framework for medical image processing and analysis via deep active learning technique. The framework includes (1) applying deep active learning approach to segment specific regions of interest (RoIs) from raw medical image by using annotated data as few as possible; (2) generative adversarial Network is employed to enhance contrast, sharpness, and brightness of segmented RoIs; (3) *Paced Transfer Learning* (PTL) strategy which means fine-tuning layers in deep neural networks from top to bottom step by step to perform medical image classification or regression tasks. In addition, in order to understand the necessity of deep-learning-based medical image processing tasks and provide clues for clinical usage, class active map (CAM) is employed in our framework to visualize the feature maps. To illustrate the effectiveness of the proposed framework, we apply our framework to the bone age assessment (BAA) task using RSNA dataset and achieve the state-of-the-art performance. Experimental results indicate that the proposed framework can be effectively applied to medical image analysis task.

1. Introduction

Recently, deep learning has achieved significant success in medical image processing and analysis. Tasks such as classification, where each medical image is assigned to a category label, are now almost exclusively done with deep learning technique.

A problem often cited when applying deep learning methods to medical image analysis is the lack of annotated training data, even if larger unlabeled data sets are more widely available [1, 2]. Manual labeling for medical images is expensive, time-consuming, and requires experienced doctors. Therefore, reducing the amount of labeled data is crucial for deep-learning-based medical image processing tasks and training a deep neural network with limited labeled data is challenging.

In general, a common framework applied on medical image classification or regression via deep learning technique contains image segmentation, image enhancement, and prediction, such as bone age assessment [3–5], pneumonia detection on chest X-rays [6], mammographic mass classification [7], diabetic retinopathy classification [8], brain tumor analysis [9], etc. All the existing works in the field of medical image processing and analysis are focused on one aspect such as segmentation, detection, and classification with an exception of [10] in which four experiments of different processing and analysis tasks in different modes are performed and the methodology of transfer learning is summarized. In this paper, we further extend the methodology and develop it into a framework for medical images processing. The proposed framework aims to address the mentioned problems and includes three key points:

(1) To alleviate human annotation burden, we employ a technique called deep active learning (AL) to actively select unlabeled samples with informative information for human to label in each training iteration.

(2) Since the medical image samples often vary considerably in intensity, contrast, and brightness, it is necessary to enhance the image quality and normalize the images for training the models of classification or regression. We use deep auto encoder network with adversarial training to tackle this problem.

(3) To positively utilize the knowledge of the source model and fine-tune parameters in medical image processing task, we propose *Paced Transfer Learning* (PTL) to fine-tune the deep convolutional neural network (CNN) according to designated rules.

The proposed versatile framework can be easily applied to different medical image classifications or regression tasks with limited annotation data and further improves the model performance. To further illustrate the effectiveness of the proposed framework, we applied the framework on bone age assessment (BAA) task. We assess the performance using the proposed method on the public dataset from 2017 pediatric bone age challenge organized by the Radiological Society of North America (RNSA) [11]. The overview of our proposed framework with application to BAA task is shown in Figure 1. The demonstrated method achieves the accuracy with mean average error (MAE) of 5.991 and 6.263 months for male and female cohorts, which achieves the state-of-the-art performance.

2. Methodology

The main contributions of this paper are as follows:

(1) We propose a method of deep AL and apply it to medical image semantic segmentation task. By using deep AL with Query By Committee (QBC) [12] strategy, we can significantly relieve manual annotation burden while the model accuracy being guaranteed.

(2) We propose a novel medical image prepossessing engine that consists of a GAN to enhance the quality of images and normalize grayscale-based medical images.

(3) We propose PTL strategy to fine-tune the off-the-shell deep CNN for specific tasks and ensure the model achieving impressive performance compared with the conventional method in deep transfer learning.

2.1. Medical Image Segmentation. Extracting specific RoI from raw medical image can significantly reduce searching space and relieve computation burden. In addition, subtracting irrelevant noise in medical image can improve model performance in the prediction stage. However, it is not easy to establish a nonlinear mapping from raw medical images to specific RoIs because of the variation of contrast, sharpness, and brightness in medical images.

Even though deep-learning-based image segmentation tasks achieve remarkable performance, a large number of annotated images are necessary. In practice, labeling work is time consuming and may need expert knowledge. The goal of active learning is to learn a classifier in a setting where data come unlabeled and any labels must be explicitly requested and paid for. The hope is that an accurate classifier can be found by buying just a few labels [13]. Under this circumstance, in order to segment specific RoIs from raw medical images, we propose deep active learning approach to alleviate annotation burden while guaranteeing model accuracy using as few labeled data as possible.

For better understanding of our image segmentation approach, it is necessary to explain the essential function of deep neural networks which is the most updated classification network and will be transferred to our domain of medical image and our task. Densely connected convolutional networks (DenseNets) [14] have shown compelling accuracy and brilliant convergence behaviors on several large-scale image recognition tasks. Meanwhile, skip connections from the down sampling to the up sampling path are usually adopted to recover spatially detailed information by reusing features maps [15]. In order to leverage the powerful capability of deep DenseNet and tackle the object efficiently and effectively, transition up and transition down blocks proposed in [16, 17] are employed as fully connected DenseNet (FC-DenseNet) to perform image semantic segmentation. The structure of FC-DenseNet which is adopted into our application is depicted in Figure 2.

We use three transition down blocks and three transition up blocks. The middle layer in whole network contains 192 feature maps, and we flatten this feature map to represent high-level feature of input X-ray image, as the purple block shown in Figure 2.

In the research field of image semantic segmentation, a pixelwise loss function is usually used to penalize the distance between the ground truth and the predicted probability map. Often, the pixelwise loss function is defined by a cross entropy as follows:

$$L_{\text{pixel\_wise}} = \sum_{\text{i}} - y_{\text{i}} \log(\hat{y}_{\text{i}}) - (1 - y_{\text{i}}) \log(1 - \hat{y}_{\text{i}}), \quad (1)$$

where y_{i} is a binary value of the corresponding pixel i and $\hat{y}_{\text{i}}$ is a predicted probability for the pixel.

DICE coefficient is another useful metric to evaluate the quality of segmentation, since it considers the overlapping between segmented result and ground truth. What is more important, in medical image segmentation, the border continuity can be improved for models with DICE loss [18]. The DICE coefficient is defined as follows:

$$\text{DICE} = 2 \frac{(|g(\hat{y}) \cap y|)}{(|g(\hat{y})| + |y|)}, \quad (2)$$

where y is a segmented mask of the corresponding image and $g(\hat{y})$ is the postprocessed binary hand mask on predicted probability map with OTSU algorithm [19].

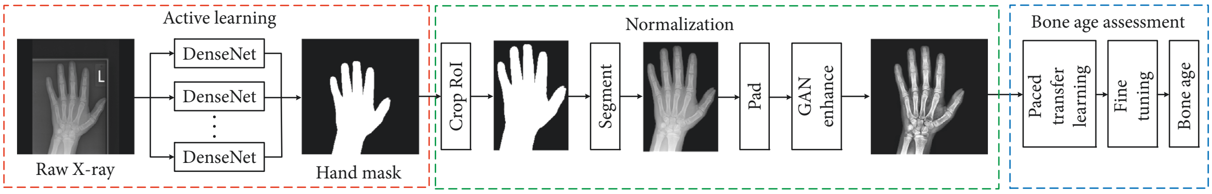


Figure 1: Overview of the proposed medical image processing framework with the application of BAA.

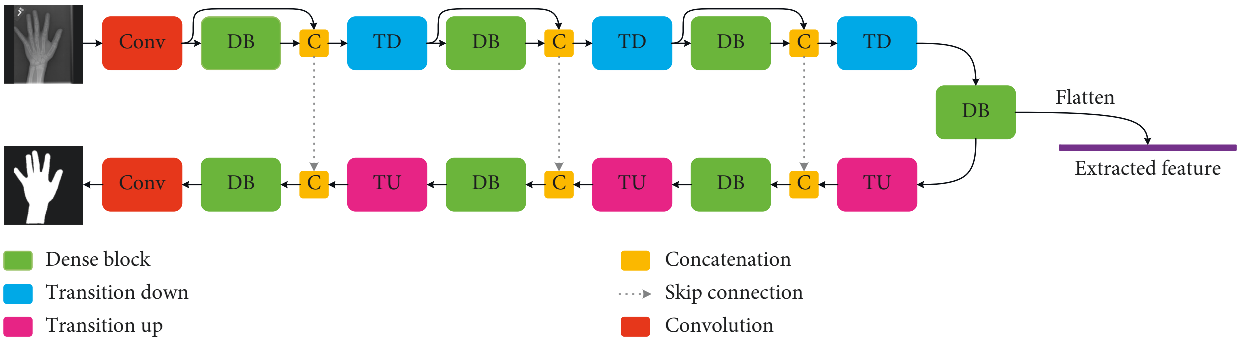


Figure 2: The architecture of refined FC-DenseNet.

High DICE coefficient contributes to powerful performance of the deep neural network and accurate segmentation result. In order to optimize neural networks by DICE coefficient, we choose $L_{\text{DICE}} = 1 - \text{DICE}$ as a penalty function and minimize it. Finally, our loss function is

$$\begin{aligned} L_{\text{loss}} &= L_{\text{pixel\_wise}} + L_{\text{DICE}},\\ &= \sum_{\text{i}} \left(-y_{\text{i}} \log\left(\hat{y}_{\text{i}}\right) - \left(1 - y_{\text{i}}\right)\log\left(1 - \hat{y}_{\text{i}}\right)\right)\\ &\quad + \left(1 - 2\frac{|g(\hat{y}) \cap y|}{|g(\hat{y})| + |y|}\right), \end{aligned} \tag{3}$$

which is equivalent to

$$\sum_{\text{i}} - y_{\text{i}} \log\left(\hat{y}_{\text{i}}\right) - \left(1 - y_{\text{i}}\right)\log\left(1 - \hat{y}_{\text{i}}\right) + \left(-2\frac{|g(\hat{y}) \cap y|}{|g(\hat{y})| + |y|}\right). \tag{4}$$

2.2. The Application of Deep Active Learning. The main hypothesis in the AL framework is that the learner can choose specific data which contain the most abundant information for oracle annotation. Recently, a lot of image segmentation methods using AL strategies have been proposed. In the vast majority of cases, active learners use uncertainty sampling strategies [20] to select unlabeled data which contain significant information to be labeled by oracle [21–24]. The key point of uncertainty sampling strategy is to measure the uncertainty of data. To address this problem, the existing algorithms often use the concepts of least confident [20], data diversity [23], cross entropy [22], etc.

In contrast, our task is to select images through QBC strategy, because QBC can take advantage of more than one model. In QBC, each member in the committee is trained on the same dataset. The next query is chosen according to the principle of maximal disagreement. In the research field of image semantic segmentation, a member in committee represents an image segmentation neural network. Formally, a committee is defined as $C = \left\{\theta^1, \theta^2, \ldots, \theta^{\text{C}}\right\}$ and $|C|$ indicates the number of committee members.

In practice, it is necessary to intellectually select data at the training stage in AL framework. This time, we use uncertainty sampling to select the most informative query instance with which the committee disagrees most. In the image segmentation task, we need to define the uncertainty of data, i.e., the disagreement level of the data. Since we trained a set of members in committee, each member learns the high-level feature of input image, as shown in the purple vector in Figure 2, and the disagreement level can be defined as sine dissimilarity:

$$\text{dissim}_{ij} = \sqrt[2]{1 - \left(\frac{\text{vector}_i \cdot \text{vector}_j}{\left|\text{vector}_i\right| \times \left|\text{vector}_j\right|}\right)^2}, \tag{5}$$

where the vector_i represents the high-level feature vector extracted by member θ^i in committee C. The dissimilarity between each member can be formulated as matrix:

$$\text{dissim} = \begin{bmatrix} 0 & \text{dissim}_{12} & \cdots & \text{dissim}_{1\text{C}} \\ \text{dissim}_{21} & 0 & & \text{dissim}_{2\text{C}} \\ & \vdots & \ddots & \vdots \\ \text{dissim}_{\text{C}1} & \text{dissim}_{\text{C}2} & \cdots & 0 \end{bmatrix}. \tag{6}$$

With equations (5) and (6), we can deduce that the dissimilarity of a single image as following:

$$\text{dissim}_{\text{img}} = \sum_{i>j>0} \text{dissim}_{ij}. \tag{7}$$

The data with highest dissimilarity indicate the most disagreement level in the committee, and the data contain the most significant information for model training. Oracle needs to give the annotation of such data as ground truth and then add the labeled data for next training epoch.

In summary, the proposed deep AL algorithm for medical image segmentation is depicted in Algorithm 1.

The proposed deep AL framework on the application of BAA is shown in Figure 3.

2.3. Image Enhancing via Generative Adversarial Network. Medical image data often vary considerably in intensity, contrast, and brightness; it is necessary to enhance the image quality and normalize the images to the classification and regression model training. To some extent, image enhancement can be defined as an image translation task where an output enhanced image is generated from an input original segmented image.

Generative adversarial network (GAN) is a generative model that creates outputs as realistic as the gold standard [25]. Usually, a GAN consists of two networks, a discriminator and a generator. The former tries to distinguish whether the image is from gold standard or outputs generated by generator, while the latter tries to generate outputs as realistic as the discriminator cannot differentiate it from the gold standard.

In our framework, we define the generator G be a map from a segmented medical image x to an enhanced image y, formally, $G : x \longrightarrow y$. The network structure of G is defined as a U-Net. The discriminator D maps a pair of $\{x, y\}$ to binary classification $\{0, 1\}$, where 0 and 1 indicate whether y is gold standard or generated by G. The network structure of D includes three CNN layers and one FC layer. The relationship between G and D on the application of BAA is shown in Figure 4.

Because the image segmentation task can be defined as an image generation task, we adopt equation (4) as the loss function of G. Then, the objective function of GAN for medical image enhancement task can be formulated as

$$\begin{aligned} L_{\mathrm{GAN}}(G, D) = {} & E_{x,y\sim p_{\mathrm{data}(x,y)}}[\log(D(x, y))] \\ & + E_{x\sim p_{\mathrm{data}(x)}}[1 - \log(D(x, G(x)))]. \end{aligned} \tag{8}$$

Note that, in conditional GAN [26], G takes a random noise to generate images, while in our task, G takes segmented medical images to generate enhanced images. Then, the optimization problem can be defined as

$$\begin{aligned} G^* = \arg\min_G \Big[& \max_D E_{x,y\sim p_{\mathrm{data}(x,y)}}[\log(D(x, y))] \\ & + E_{x\sim p_{\mathrm{data}(x)}}[1 - \log(D(x, G(x)))] \Big]. \end{aligned} \tag{9}$$

For D, the goal is to correctly distinguish whether the image is a generated or a gold standard; the optimization objective for D is

$$\begin{aligned} D^* = \arg\max_D & E_{x,y\sim p_{\mathrm{data}(x,y)}}[\log(D(x, y))] \\ & + E_{x\sim p_{\mathrm{data}(x)}}[1 - \log(D(x, G(x)))]. \end{aligned} \tag{10}$$

2.4. Paced Transfer Learning for Medical Image Classification or Regression. It is demonstrated in [10] that training a deep CNN from scratch is difficult because it requires a large amount of labeled training data. Fortunately, a promising alternative is to fine-tune a pretrained CNN which could outperform a CNN trained from scratch. In our common sense, the lower layers of a CNN learn low-level image features, such as shape, edge, etc., while the higher layers learn high-level features, which are more important to specific application.

In general, the neural network for transfer learning contains one off-the-shell CNN followed by several fully connected (FC) layers. The weights in the off-the-shell CNN are initialized by the pretrained weights in source field; however, the parameters in FC layers are sampled from a normal distribution with a zero mean and small standard deviation. It is illustrated in [10] that the stochastically initialized weights often cause large noise, and when we optimize CNN through gradient decent styled optimization algorithm, the fine-grained parameters will jump out of the global optimal solution and may lead to an undesirable local minimum. In this situation, if we fine-tune all the layers at the initial stage, the well-trained weights may be overwritten. What is more severely, the solution may not go back to the optimal solution because we only have a limited amount of medical image data for model training.

To address this problem, we proposed the *Paced Transfer Learning (PTL)*. PTL means that fine-tuning layers in deep neural network from top to bottom step by step. At the initial stage, we only fine-tune the random initialized FC layers in the top of neural networks. Then, as the loss decreases to a stable state, we further fine-tune the second top layers. With several fine-tuning steps, all layers in the deep neural network are fine-tuned together until the model converges.

To illustrate how PTL works, here we use Xception and apply it to the task of BAA to explain the detail of the fine-tune process. Due to the size of feature maps, naturally, Xception can be split into three blocks: Entry flow block, middle flow block, and exit flow block, as depicted in Figure 5. By using the proposed PTL, we sequentially fine-tune parameters in each block from top to bottom. At the first step, we only train the parameters in FC layers while fixing other parameters. As the loss converges to a stable state, we fine-tune the parameters in the blue rectangle. Before training one of the blocks, the fine-tuned parameters in previous blocks have to be finished training for specific task. Therefore, PTL prevents from overwriting the fine-tuned parameters and makes it possible to achieve a positive refinement of the adopted off-the-shell CNN.

2.5. Instruct Clinical Practice via Class Activation Map. Despite CNN achieves impressive performance, it is necessary to investigate the essential function of CNN and provide explanations to clinical practice. The higher-level layers of CNN, such as the FC layer in the top of CNN, represent very effective generic features for image recognition task. As is demonstrated in [15], a class activation map

```
Input:
$L = \{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$: initial labeled training data, composed of $n$ samples;
$U = \{\{x_1\}, \{x_2\}, \ldots, \{x_m\}\}$: initial unlabeled training data, composed of $m$ samples;
$C = \{\theta^1, \theta^2, \ldots, \theta^C\}$: committee of medical image segmentation networks to be trained;
Output:
$C = \{\theta^1, \theta^2, \ldots, \theta^C\}$: trained committee of medical image segmentation networks
Repeat:
    1. Train $C = \{\theta^1, \theta^2, \ldots, \theta^C\}$ with the loss function in equation (4) on the labeled data.
    2. Calculate the dissimilarity of each sample in $U$ among every member in C and select the data with the larger dissimilarity.
    3. Oracle is queried to annotate the data selected in step 2 and add the annotated sample to $L$.
    4. Update $U$ and $L$.
Until:
The $C = \{\theta^1, \theta^2, \ldots, \theta^C\}$ is converged to a satisfied result.
```

Algorithm 1: Proposed deep AL framework.

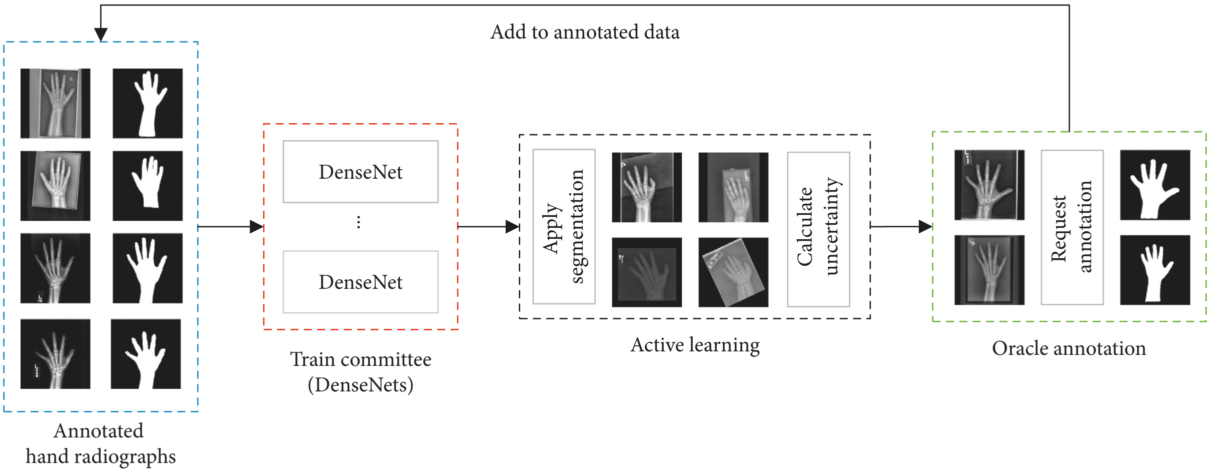


Figure 3: The overview of deep AL framework for medical image semantic segmentation on the application of BAA.

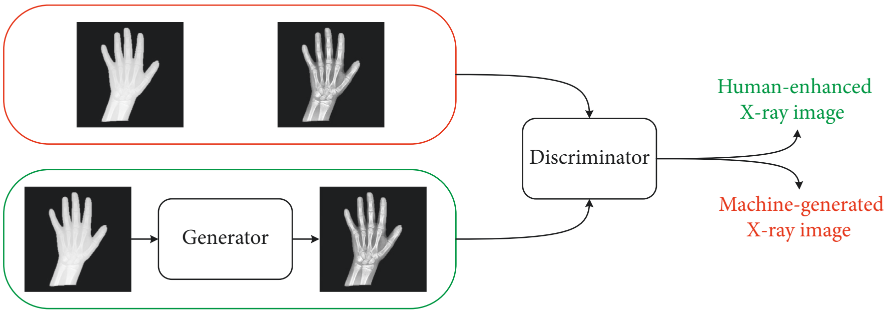


Figure 4: GAN for medical image enhancement.

(CAM) for a particular category indicates the discriminative image regions used by the CNN to identify that category. In the proposed network, the FC layer in the top of CNN followed with a global average pooling (GAP) layer. The essential function of the GAP layer is outputting the spatial average of the feature map of each unit at the last CNN layer, and FC layer uses the weighted sum operation to generate the final output. Hence, the GAP layer and FC layer reflect which part of feature map is crucial for the final output. CAM maps the predicted class score back to the previous CNN layer to generate the RoI which is crucial for model training.

3. Application and Results

To illustrate the effectiveness and performance of the proposed framework, we apply our framework to the task of BAA on a public available dataset from RSNA Bone Age

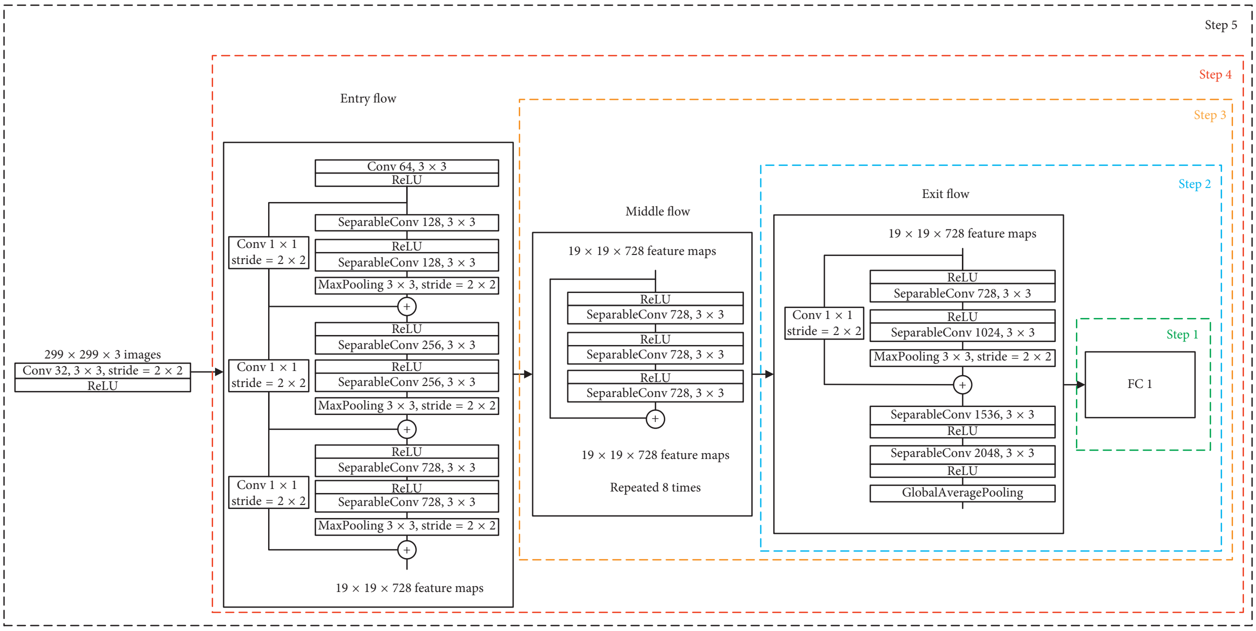


Figure 5: PTL on Xception. Parameters in different blocks are fine-tuned in designated steps.

Challenge [11]. Skeletal BAA is a common clinical practice for evaluating the stage of skeletal maturation of a child [17, 27]. An incompatibility between the chronological age and development of bone age indicates abnormalities in skeletal development. Radiologists could evaluate the growth disorder, monitor the hormone therapy, and predict adult height through BAA. To perform BAA, using our framework, we need to segment hand RoI from raw X-ray radiographs, enhance the quality of hand RoI, and use transfer learning to predict bone age.

3.1. Data Overview. The hand radiographs provided by RSNA contain 12611 cases in training dataset. The bone age ranges from 1 month to 228 months as shown in Figure 6.

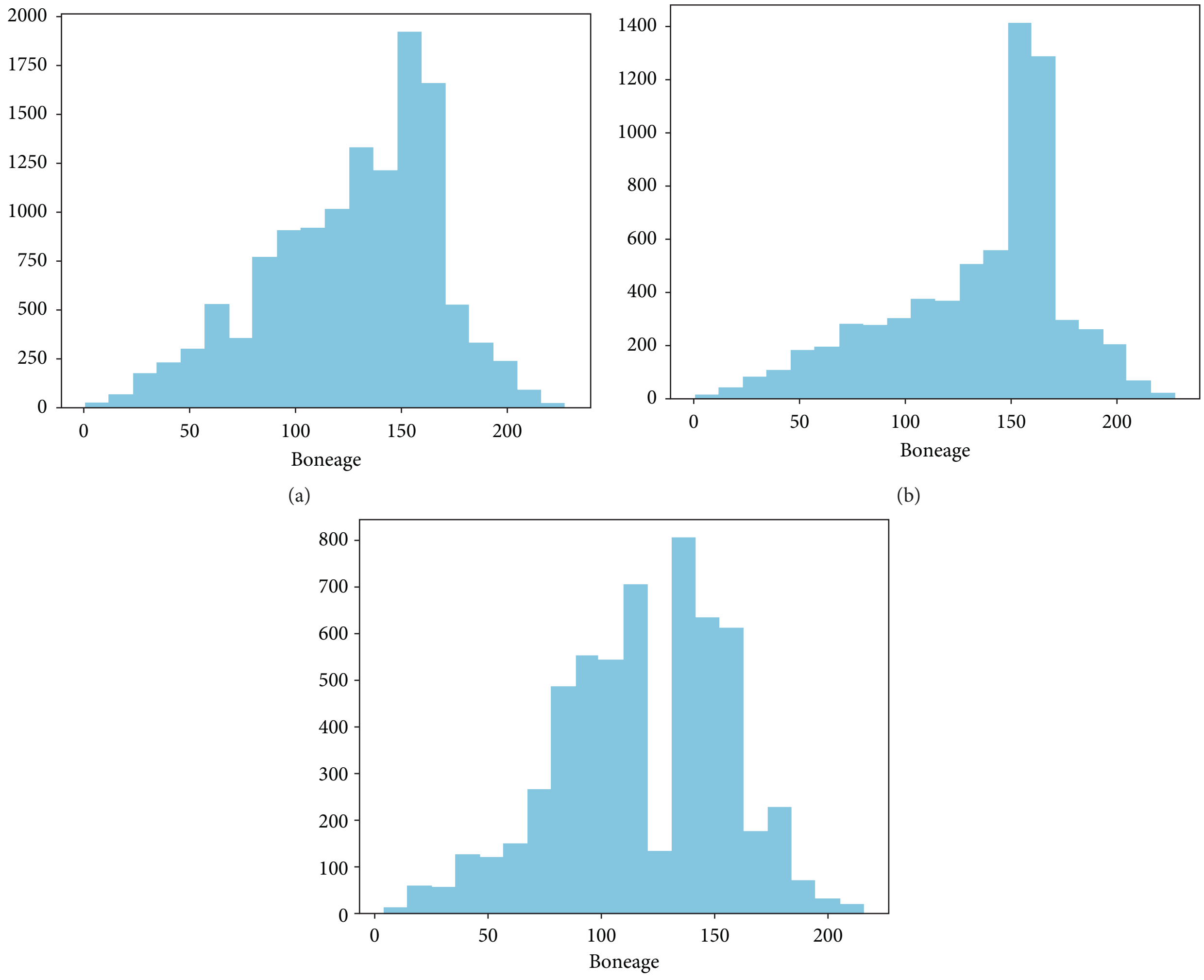
The hand radiographs have diverse shape, grayscale, and size. Some images have black bones with white background and vice versa. Worse still, the images are randomly rotated with different angles and the borders of hands are indistinct in some instances. Figure 7 shows an illustrating view of raw X-ray images.

3.2. Deep Active Learning for Hand Segmentation. In this section, we use FC-DenseNet as hand radiographs segmentation network and employ QBC strategy to actively segment hand RoI from raw X-ray images. The detail of the training method is illustrated in Algorithm 1. In our experiments, we set the committee size of three. We train three FC-DenseNets with the architecture depicted in Figure 2 and initialize the model parameters with different random seeds so that the feature vectors are different from each other. In each training iteration, the proposed deep AL framework selects most informative data in unlabeled dataset and asks human oracle for hand masks.

At the training initialization stage, we manually annotate 100 raw X-ray images. The designed interactive program could tell us which data are crucial for training at each training iteration, then we give the ground truth of the X-ray images, and add them to labeled training dataset. To determine whether the model is converged, we use the trained DenseNets to infer the unlabeled data and visually inspect the segmentation results. Totally, we annotate 400 raw X-ray images, and the refined FC-DenseNets with AL can precisely segment all of data. Some images selected by AL at several epochs are shown in Figure 8. We observed that the selected raw images with most dissimilar features have obvious different segmentation results. This phenomenon suggests the proposed AL framework with QBC strategy works effectively. And the segmentation results are shown in Figure 9. After segmenting hand masks, we cropped the hand RoIs from full image, as shown in the last line in Figure 9.

As a comparison, we use U-Net to train hand segmentation network with only 400 annotated hand RoI without AL technique. The segmentation results are shown in the second row in Figure 9. It is clear that with AL training trick, the segmentation performance is significantly improved. This indicates the proposed AL algorithm could effectively select data for segmentation network training and generate segmentation results with high quality.

3.3. Hand RoI Enhancement via Generative Adversarial Network. In our experiment, we manually adjust contrast, brightness, and sharpness of 500 hand segmentation images and denote the preprocessed image as gold standard. The manual annotated images and corresponding original images constitute the training data. To prevent overfitting and improve model performance, we augment data by flipping image around vertical axes and randomly rotating from −30°



(a)

(b)

(c)

Figure 6: Bone age distribution for all (a), female (b), and (c) male left hands.

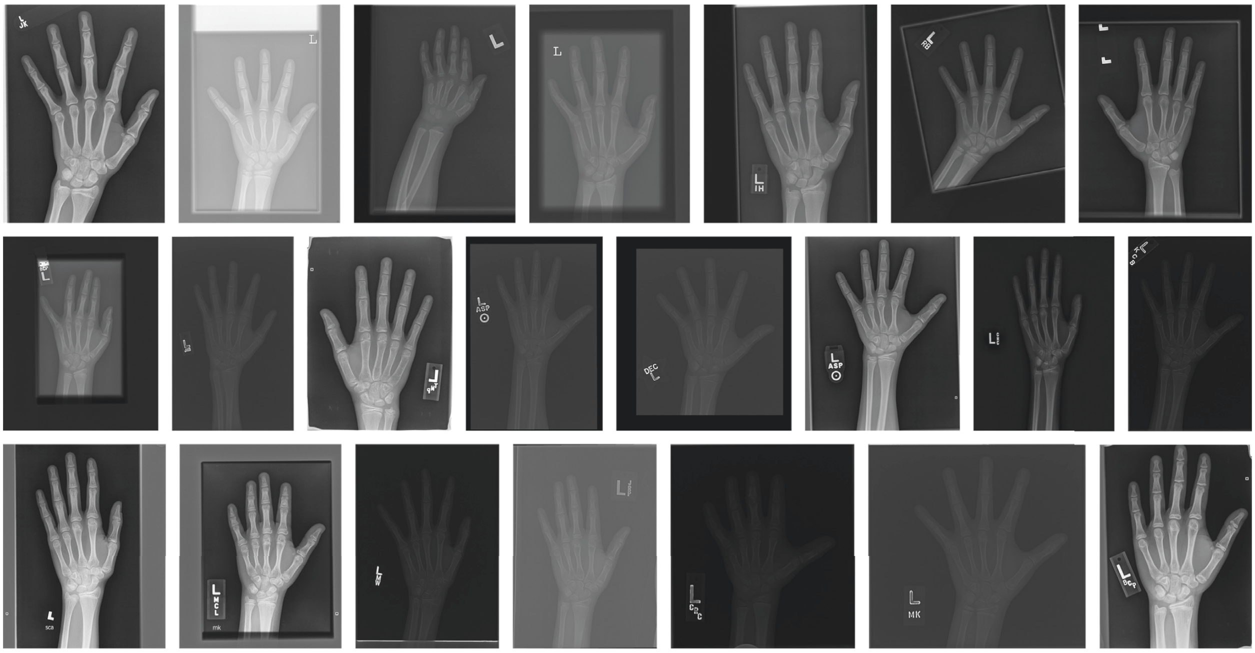


Figure 7: Some troublesome examples in RSNA hand bone dataset.

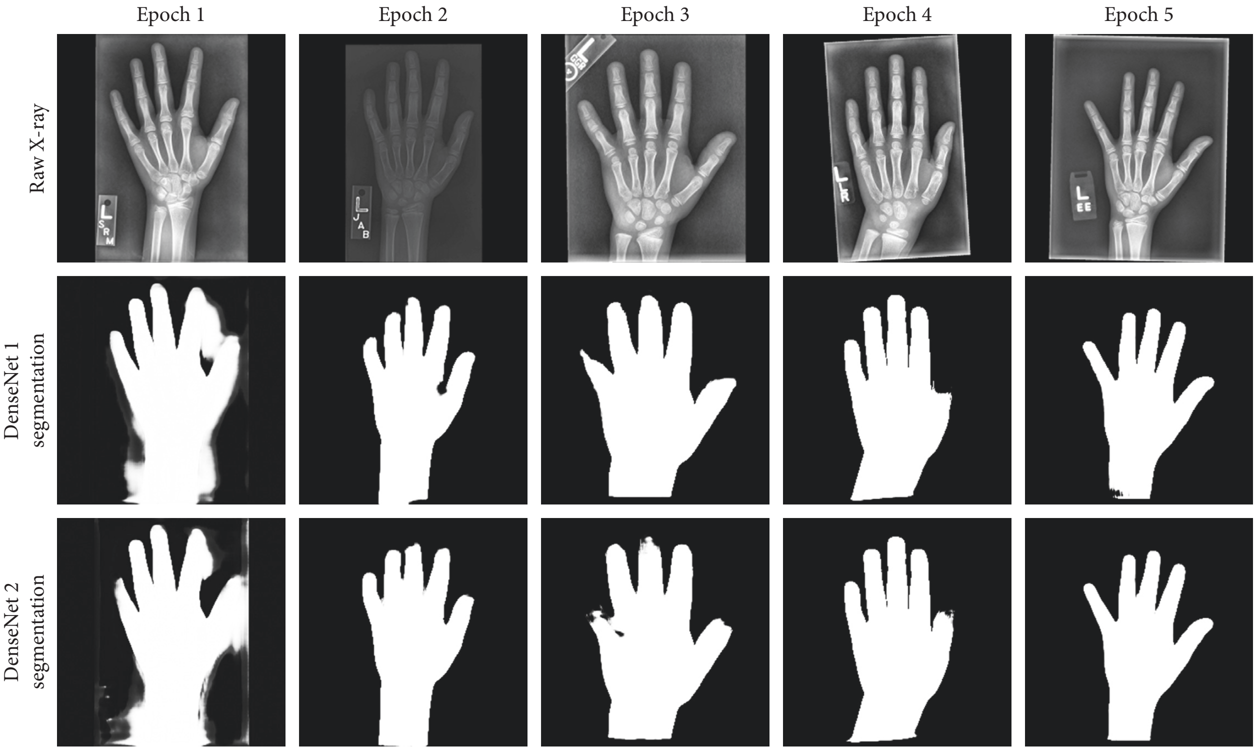


Figure 8: Examples to show how AL selected the most uncertain image to annotate.

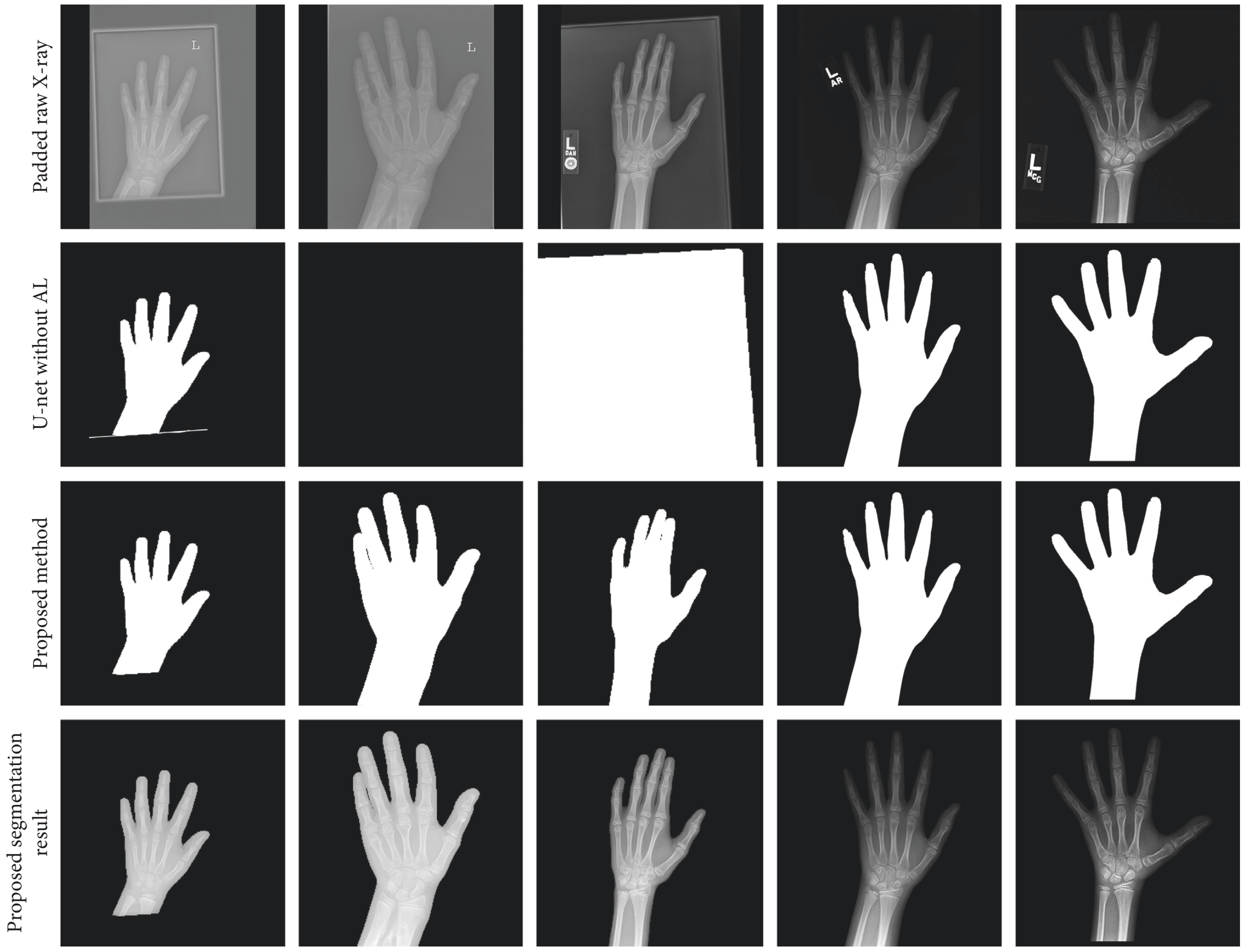


Figure 9: Segmentation results of hand radiographs.

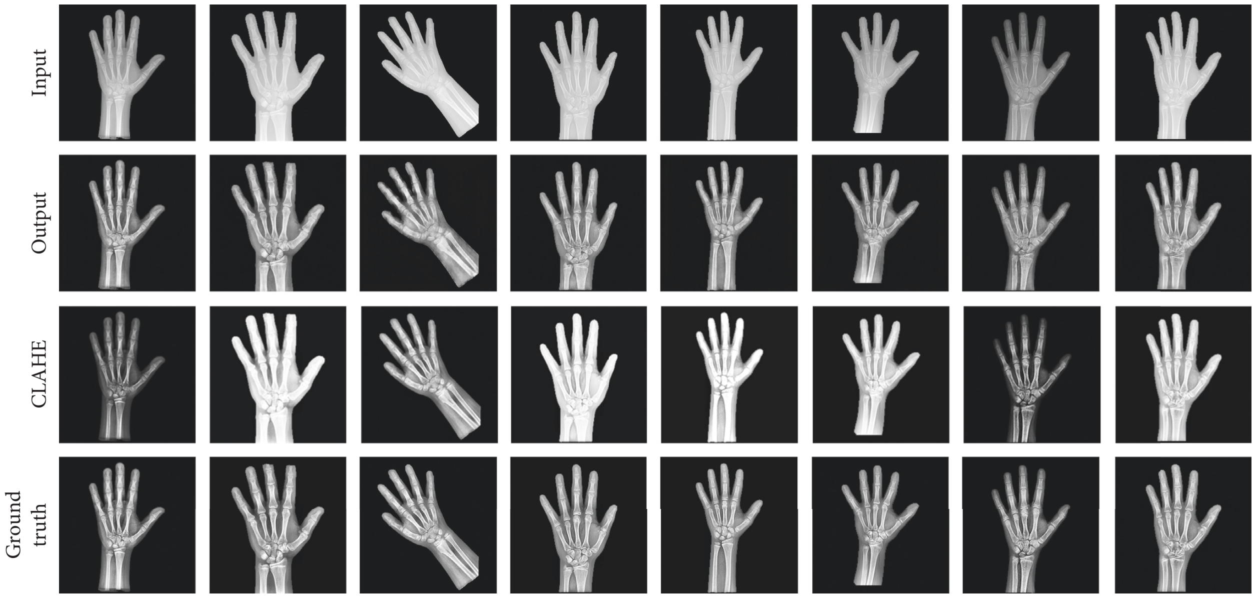


Figure 10: Image enhancement results. From top to bottom are original segmented hand image, enhanced image, and ground truth (gold standard).

to 30°. The image enhancement results are shown in Figure 10.

From Figure 10, we can observe that enhanced images have a strong resemblance of gold standard, and it is hard to distinguish gold standard from generated image. As a comparison, we use contrast-limited adaptive histogram equalization (CLAHE) to process the segmented images, as shown in the third row in Figure 10. It is clear that the quality of images processed by CLAHE is less stable than GAN generated images. This suggests the G has powerful ability to reliably enhance all images in dataset.

3.4. Bone Age Assessment via Paced Transfer Learning. To prevent overfitting, we adopt online data augmentation with random rotation, random zoom, random sheer, and horizontal flipping. Furthermore, we sample a same number of data in each bone age to keep data category balance. We implemented our models with Keras and TensorFlow, and we trained the models on a NVIDIA-DGX-1V machine equipped with 8 Tesla V100 GPUs. In addition, we trained the model through RMSprop algorithm with 0.001 of the base learning rate, and the learning rate decreased with decay of 0.05. Each training epoch only took 3 minutes. Our source code is available in https://github.com/awp4211/bone-age-assessment.

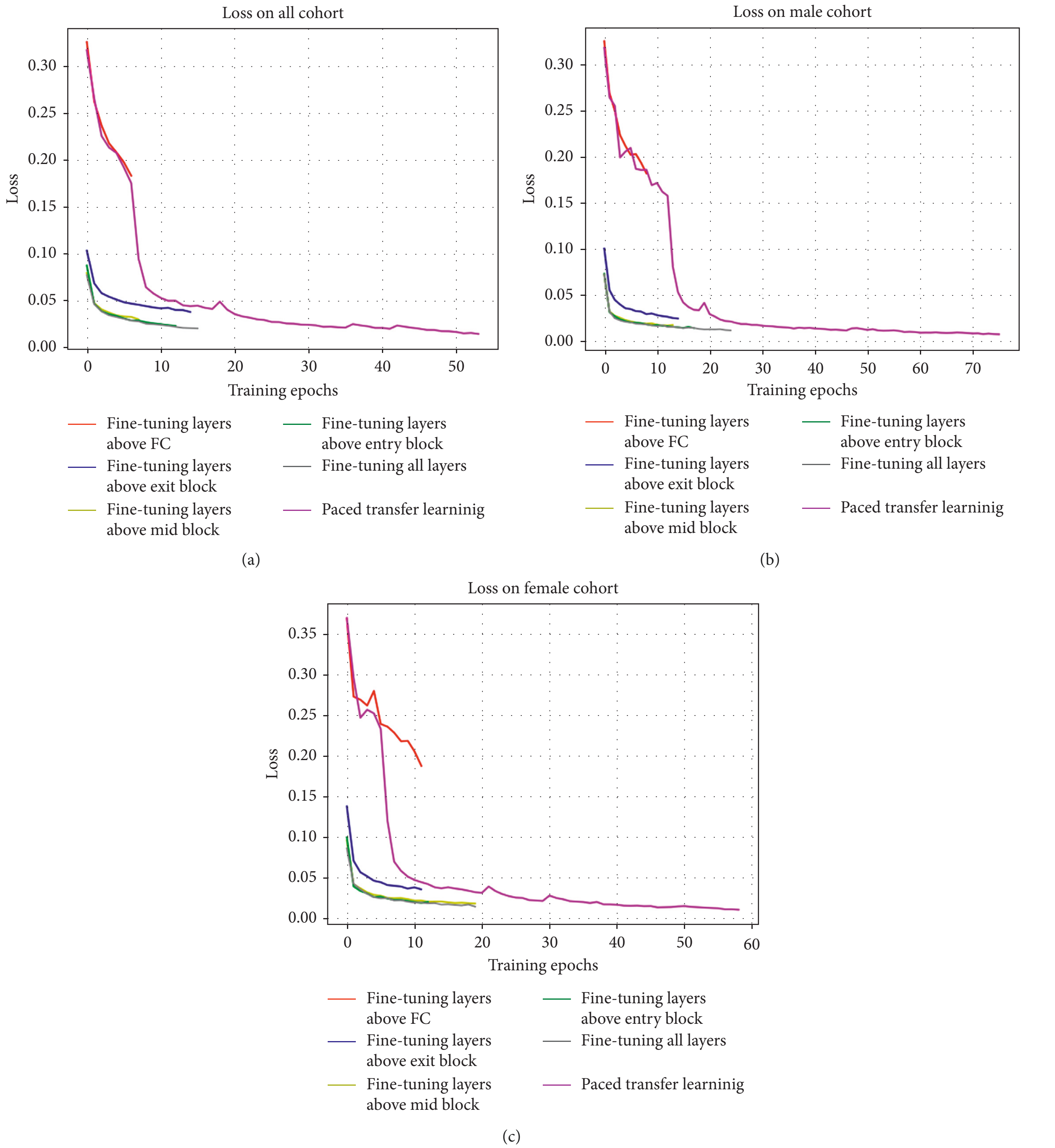
To prove the proposed paced transfer learning strategy outperforms conventional transfer learning methods, we monotonously fine-tuned the parameters in each block in Figure 5 while we fixed the parameters outside of the block. Figure 11 presents the loss for different fine-tune model.

Each transfer learning model was trained with enhanced hand radiographs which are resized to 299 × 299. In each part of dataset, all, male, and female cohorts, our model achieves best performance with proposed PTL technique compared to other transfer learning settings. A further conclusion is that even though we fine-tune a larger quantity of parameters in whole CNN at beginning, i.e., fine-tuning from entry block and fine-tuning all layers, model loss did not decrease to a lower level. This suggests fine-tuning all layers at the initial stage might deteriorate fine-grained parameters. On the contrary, fine-tuning a small number of parameters is insufficient for model training. This circumstance might occur in transfer learning on medical image processing tasks because there exists a huge difference between nature images and medical images. Performance of models with different transfer learning strategies is compared in Figure 12. Fine-tuning with paced transfer learning surpasses others significantly.

From Figure 12, we concluded that paced transfer learning can remarkably improve model performance and reduce MAE. By observing the first three columns in Figure 12, we conclude that, by using enhanced hand segmentation images, our model gains a better performance. The proposed BAA model achieved a MAE of 7.664, 5.991, and 6.263 months on all, male, and female cohorts, respectively. More importantly, separating patients into male and female cohorts could slightly improve model performance. The reason behind this phenomenon is that male and female cohorts are judged with different standards in clinical practice [28].

In addition, we compare our method with existing BAA with deep learning approaches and summarize the MAE and dataset size in Table 1. The proposed method acquired a better performance on a large-scale dataset which contains 12611 cases.

By observing Table 1, it is clear that our proposed model dominates over other methods in part of MAE evaluation and training time. Compared with two previous BAA approach through transfer learning in [4, 5], we could conclude that the proposed PTL technique significantly transfers



(a)

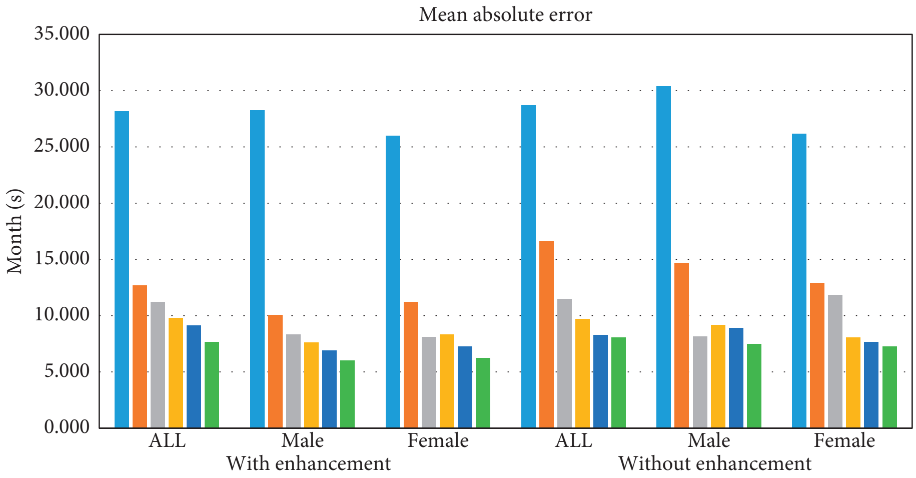
(b)

(c)

Figure 11: Raining loss for fine-tuning from different blocks on all, male, and female cohorts. Since that, we used early stopping technique, and different models were trained with different epochs.

pretrained weights to medical image process tasks and accelerates the training procedure. The proposed PTL successfully prevent trapping into local minimum and makes model converge at several training steps. In addition, compared with the two approach which trained in a relative small dataset in [3, 29], our model successfully leveraged all information provided in dataset and achieved a higher performance.

What is more important, due to the use of PTL technique, our model converged within 75 epochs in each cohort of dataset and the training time was much less. In contrast to all previous approaches in Table 1, we use online sampling for data augmentation to successfully prevent overfitting and each training epoch is only takes 3 minutes.



	With enhancement			Without enhancement		
	ALL	Male	Female	ALL	Male	Female
🟦 Fine-tuned layers above FC	28.168	28.271	25.982	28.731	30.388	26.180
🟧 Fine-tuned layers above exit block	12.695	10.082	11.242	16.635	14.702	12.915
⬜ Fine-tuned layers above mid block	11.213	8.339	8.115	11.511	8.146	11.835
🟨 Fine-tuned layers above entry block	9.796	7.626	8.337	9.695	9.175	8.060
🟦 Fine-tuned all layers	9.141	6.923	7.265	8.305	8.923	7.666
🟩 Paced transfer learning	7.664	5.991	6.263	8.074	7.495	7.256

Figure 12: Comparison of models with different transfer learning strategies with respect to MAE (months).

Table 1: Comparison of different deep-learning-based methods on BAA task with respect to MAE and dataset size.

Method	Feature	Dataset size	MAE (m)	Training epoch
Zhou et al. [29]	Transfer learning	1100	8.63	40
Spampinato et al. [3]	BoNet (CNN)	1400	9.48	150
Lee et al. [5]	Transfer learning	9325	11.16 (M)/9.84 (F)	100
Iglovikov et al. [4]	Transfer learning	12611	6.30 (M)/6.10 (F)	—
Proposed method	PTL	12611	5.991 (M)/6.263 (F)	<75

Table 2: Comparison of model performance with the participants in RSNA competition.

Rank	User	MAE
1	Elmigu	5.796
2	Jeffmenin	5.830
3	Bratta	5.911
4	Alexandrecadrin	6.102
5	s8t	6.123
6	Felipe.kitamura	6.164
7	S.Koitka	6.180
8	Leonchen	6.209
9	Jcrayan	6.288
10	Elmigu	6.365
-	Proposed	5.991 (M)/6.263 (F)

To further illustrate the performance of the proposed approach for BAA, we compare our methods with the leader board available on RSNA challenge website.

From Table 2, we can find that our results achieve rank 4 and rank 9 on male and female cohorts.

3.5. Visualizing CNN by Class Activation Map. Representative CAM was generated for male and female skeletal development, as shown in Figure 13.

In Figure 12, highlighted RoIs are important portions of the image to perform final bone age estimation. Compared with metacarpal bones, CAM focuses less attention on carpal bones in male skeleton, implying that metacarpal bones are important for predicting bone age for male in our method. While for female, CAM focuses on a large range of hand bones, including the tail of phalanges, metacarpal bones, and carpal bones. However, carpal bones are more crucial for determining bone ages.

4. Conclusion

In this paper, we proposed a versatile framework for medical image processing and analysis using deep learning technique. At the data preprocessing stage, we propose a deep AL framework to actively select the most informative data for annotation and segment specific RoI with FC-DenseNet. In addition, GAN is employed to enhance medical image quality. For the medical image regression or classification task, we propose PTL to fine-tune an off-the-shell CNN and perform the predication. Furthermore, we visualize the deep CNN by using CAM and explain which portions of image are crucial for computer-aided system for specific medical image processing tasks. To exemplify the effectiveness and performance of the proposed

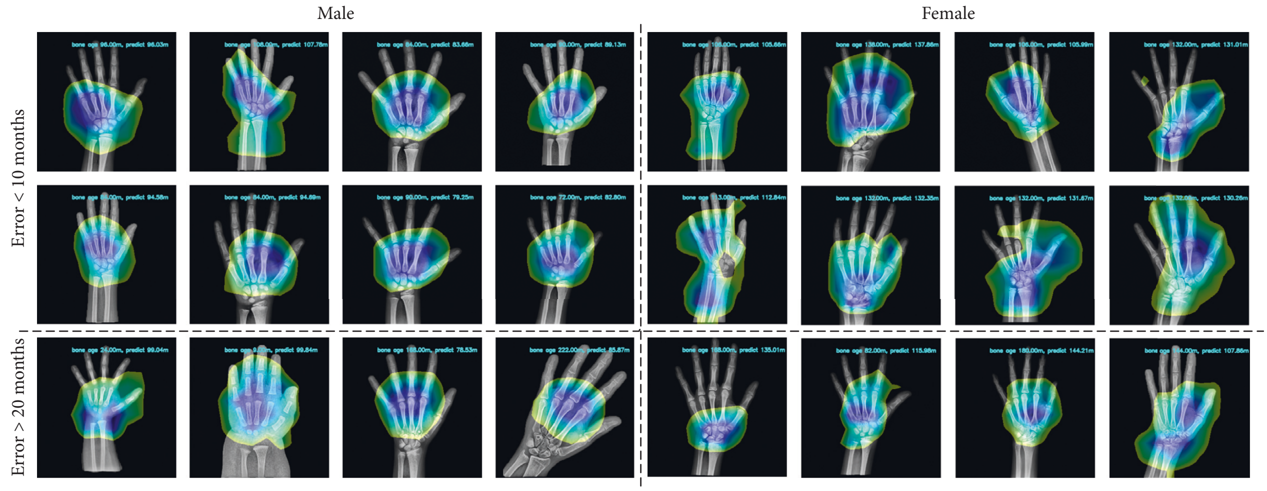


Figure 13: Representative examples of CAM for male (a) and female (b) cohorts with different discrepancies between real bone age and predicated bone age.

framework, we test our model on BAA task on RSNA dataset. Our model achieves a MAE of 5.991 and 6.263 months on male and female cohorts, comparable to the state-of-the-art performance on a large-scale dataset. We believe the proposed framework could also be applied to other medical image recognization tasks.

5. Future Work

Although the proposed approach achieved the state-of-the-art performance, there are also some limitations. One problem is that we need to annotate several images with specific RoI at the initial training stage. Due to the limited number of training samples, the model may trap into local minimum. And another limitation is the proposed approach could not fully automatically enhance the quality of medical images.

The investigation presented in this paper leaves many open challenges and issues for future research. We concisely discuss some of them in the following:

(1) Apply proposed framework on different modalities of medical images, such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound (US).

(2) Integrate different source of off-the-shell CNN to the framework, i.e., transferring other off-the-shell CNNs to medical applications.

(3) Apply AL and PTL on different types of medical image processing tasks such as detection and localization.

Data Availability

The X-Ray imaging data used to support the findings of this study have been deposited in the RSNA repository http://rsnachallenges.cloudapp.net/competitions/4.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Graduate Innovation Foundation in Xi'an University of Posts and Communications under Grant CXJJ2017005.

References

[1] V. Cheplygina, M. D. Bruijne, and J. P. W. Pluim, "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," 2018, https://arxiv.org/abs/1804.06353.

[2] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.

[3] C. Spampinato, S. Palazzo, D. Giordano et al., "Deep learning for automated skeletal bone age assessment in X-ray images ☆," *Medical Image Analysis*, vol. 36, pp. 41–51, 2016.

[4] V. Iglovikov, A. Rakhlin, A. Kalinin et al., *Pediatric Bone Age Assessment Using Deep Convolutional Neural Networks, [12] Bone Age Recognition Using Convolution Neural Network*, 2017, https://arxiv.org/abs/1712.05053.

[5] H. Lee, S. Tajmir, J. Lee et al., "Fully automated deep learning system for bone age assessment," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 427–441, 2017.

[6] P. Rajpurkar, J. Irvin, K. Zhu et al., "CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, https://arxiv.org/abs/1711.05225.

[7] T. Kooi, G. Litjens, G. B. Van et al., "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303–312, 2017.

[8] M. J. J. P. V. Grinsven, B. V. Ginneken, C. B. Hoyng et al., "Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.

[9] M. Ghafoorian, N. Karssemeijer, T. Heskes et al., "Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin," *Neuroimage: Clinical*, vol. 14, pp. 391–399, 2017.

[10] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or

fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[11] RSNA, "RSNA pediatric bone age challenge," December 2017, http://rsnachallenges.cloudapp.net/competitions/4.

[12] H. S. Seung Opper, "Query by committee," in *Proceedings of Fifth Workshop on Computational Learning Theory*, pp. 287–294, Pittsburgh, PA, USA, July 1992.

[13] S. Dasgupta, "Coarse sample complexity bounds for active learning," *Neural Information Processing Systems*, pp. 235–242, 2005.

[14] G. Huang, Z. Liu, L. V. D. Maaten et al., "Densely connected convolutional networks," 2016, https://arxiv.org/abs/1411.1784.

[15] B. Zhou, A. Khosla, A. Lapedriza et al., "Learning deep features for discriminative localization, computer vision and pattern recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, Las Vegas, NV, USA, June 2016.

[16] S. Jegou, M. Drozdzal, D. Vazquez et al., "The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1175–1183, Las Vegas, NV, USA, June 2017.

[17] D. Franklin, "Forensic age estimation in human skeletal remains: current concepts and future directions," *Legal Medicine*, vol. 12, no. 1, pp. 1–7, 2010.

[18] M. Drozdzal, E. Vorontsov, G. Chartrand et al., "The importance of skip connections in biomedical image segmentation," in *Lecture Notes in Computer Science*, pp. 179–187, Springer, Berlin, Germany, 2016.

[19] N. Otsu, "Threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–6s6, 1979.

[20] D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Proceedings of the International Conference on Machine Learning (ICML)*, New Brunswick, NJ, USA, July 1994.

[21] J. Kong, F. Wang, G. Teodoro et al., "Automated cell segmentation with 3D fluorescence microscopy images," in *Proceedings of IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 1212–1215, Brooklyn, NY, USA, April 2015.

[22] D. Mahapatra, P. J. Schüffler, J. A. W. Tielbeek et al., "Active learning based segmentation of Crohn's disease using principles of visual saliency," in *Proceedings of IEEE, International Symposium on Biomedical Imaging (ISBI)*, pp. 226–229, Beijing, China, April 2014.

[23] S. D. Jain and K. Grauman, "Active image segmentation propagation, computer vision and pattern recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2864–2873, Las Vegas, NV, USA, June 2016.

[24] D. Mahapatra and J. M. Buhmann, "Visual saliency-based active learning for prostate magnetic resonance imaging segmentation," *Journal of Medical Imaging*, vol. 3, no. 1, article 014003, 2016.

[25] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., *Generative Adversarial Nets, International Conference on Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, 2014.

[26] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *Computer Science*, pp. 2672–2680, 2014.

[27] V. Gilsanz and O. Ratib, *Hand Bone Age: A Digital Atlas of Skeletal Maturity*, Springer Nature, Basingstoke, UK, 2005.

[28] J. O. Forfar, "Assessment of skeletal maturity and prediction of adult height," *American Journal of Human Biology*, vol. 14, no. 6, pp. 788-789, 2010.

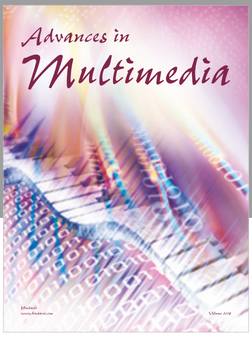
[29] J. Zhou, Z. Li, W. Zhi et al., "Using convolutional neural networks and transfer learning for bone age classification," in *Proceedings of International Conference on Digital Image Computing: Techniques and Applications*, pp. 1–6, Sydney, Australia, November 2017.

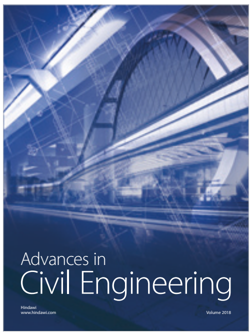
Journal of
Engineering

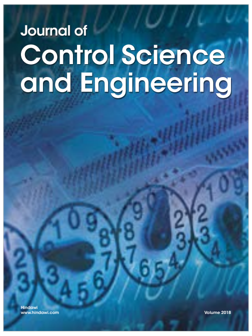
The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

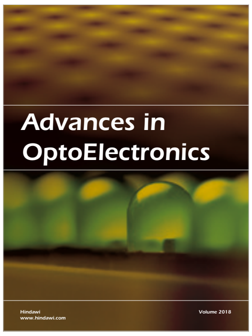

Advances in
Multimedia


Advances in
Civil Engineering


Journal of
Control Science
and Engineering

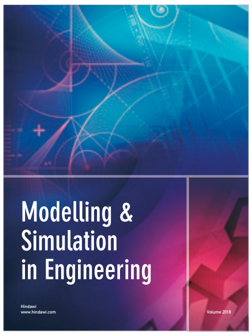
Journal of
Robotics

Journal of
Electrical and Computer
Engineering

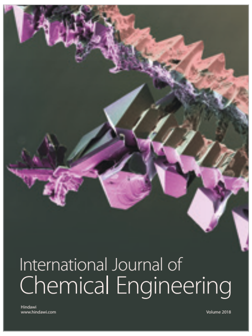

Advances in
OptoElectronics

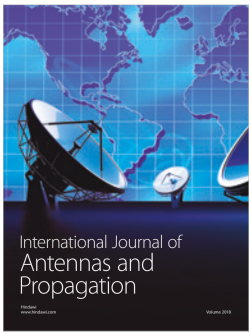
VLSI Design

International Journal of
Navigation and
Observation

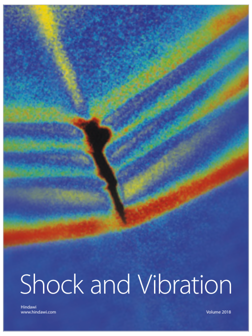

Modelling &
Simulation
in Engineering

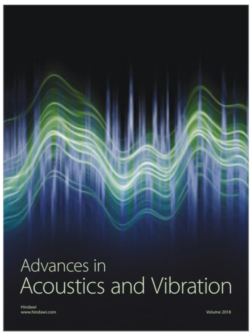
International Journal of
Aerospace
Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration