# A Siamese Network with Adaptive Gated Feature Fusion for Individual Knee OA Features Grades Prediction

**Kang Wang** ( wangkang@nudt.edu.cn )
National Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology

**Xin Niu**
National Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology

**Yong Dou**
National Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology

**Dongxing Xie**
Department of Orthopaedics, Xiangya Hospital, Central South University

**Tuo Yang**
Department of Health Management Center, Xiangya Hospital, Central South University

**Research Article**

**Keywords:** Siamese Network, Osteoarthritis (OA), knee osteoarthritis (OA) features, Adaptive Gated Feature, Grades Prediction

**Posted Date:** March 1st, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-257250/v1



**Version of Record:** A version of this preprint was published at Scientific Reports on August 19th, 2021.
See the published version at https://doi.org/10.1038/s41598-021-96240-8.

# A Siamese Network with Adaptive Gated Feature Fusion for Individual Knee OA Features Grades Prediction

**Kang Wang[1,*], Xin Niu[1], Yong Dou[1], Dongxing Xie[2], and Tuo Yang[3]**

[1]National Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology, Changsha 410073, China
[2]Department of Orthopaedics, Xiangya Hospital, Central South University, Changsha 410008, China
[3]Department of Health Management Center, Xiangya Hospital, Central South University, Changsha 410008, China
[*]wangkang@nudt.edu.cn

## ABSTRACT

Grading individual knee osteoarthritis (OA) features is a fine-grained knee OA severity assessment. Existing methods ignore following problems: (1) more accurately located knee joints benefit subsequent grades prediction; (2) they do not consider knee joints' symmetry and semantic information, which help to improve grades prediction performance. To this end, we propose a SE-ResNext50-32x4d-based Siamese network with adaptive gated feature fusion method to simultaneously assess eight tasks. In our method, two cascaded small convolution neural networks are designed to locate more accurate knee joints. Detected knee joints are further cropped and split into left and right patches via their symmetry, which are fed into SE-ResNext50-32x4d-based Siamese network with shared weights, extracting more detailed knee features. The adaptive gated feature fusion method is used to capture richer semantic information for better feature representation here. Meanwhile, knee OA/non-knee OA classification task is added, which helps extract richer features. We specially introduce a new evaluation metric (top$\pm$1 accuracy) aiming to measure model performance with ambiguous data labels. Our model is evaluated on two public datasets: OAI and MOST datasets, achieving the state-of-the-art results comparing to competing approaches. It has the potential to be a tool to assist clinical decision making.

## Introduction

Knee osteoarthritis (OA)[1] is a degenerative joint disease, the major pathological features of which are osteophytes formation and knee joint space narrowing[2–4]. The severe case may cause excruciating pain and even total joint replacement[5]. The huge expense of knee treatment is surprising, which even reaches 19000 euros one year for each patient[6]. Thus, early diagnosis and treatment are necessary for the defense of knee OA. As is well known, the growth of the computer applications has achieved great success in medical engineering, so is for knee OA diagnosis. Computer-aided diagnosis[3] reduces the subjectivity of assessing knee OA and achieves automatic knee OA diagnosis rapidly. At present, the common computer-aided diagnosis is based on radiography (X-ray)[7], which is a cheap and widely used medical imaging compared with other tools[8,9], such as magnetic resonance imaging (MRI), ultrasound imaging, etc. The gold standard of predicting knee OA severity in X-ray is Kellgren-Lawrence (KL) grading system[10], which includes KL0 (no OA), KL1 (Doubtful OA), KL2 (Minimal OA), KL3 (Moderate OA) and KL4 (Severe OA). However, KL grade is a composite score, which does not separately focus on individual features and lateral OA side/ medial OA side. Recently, Osteoarthritis Research Society International (OARSI) atlas[11] has been a feature-specific approach to grade knee OA severity. The specific features are joint space narrowing (JSN), femoral osteophytes (FO) and tibial osteophytes (TO) (see Figure 1), where JSN includes the lateral joint space narrowing (JSN-L) and the medial joint space narrowing (JSN-M), FO includes the femoral lateral osteophytes (FL) and the femoral medial osteophytes (FM), and TO includes the tibial lateral osteophytes (TL) and the tibial medial osteophytes (TM). The grades of JSN-L, JSN-M, FL, FM, TL, TM all include four grades from Grade 0 to Grade 3. This provides a fine-grained knee OA severity assessment, which plays an important role in supporting clinical decisions.

Nowadays, several studies have demonstrated success in diagnosing knee OA KL-grade from X-ray, but just a few studies about assessing individual knee OA features from X-ray. They usually locate knee joints and then make subsequent diagnosis. Tiulpin et al.[12] presented HOG and SVM method[13] to detect knee joints and a 7-layer Siamese convolutional neural network to classify knee OA KL-grade, which has good performance in KL-grade prediction. But the knee joint detection method via HOG and SVM is a traditional machine learning method, which is inferior to deep learning methods in the features extraction. In addition, Tiulpin et al. only studied knee OA KL-grade. In a later work[14], Tiulpin et al. leveraged an ensemble model of

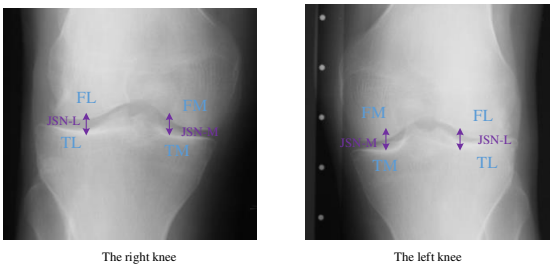



**Figure 1.** The specific features in knee images.

SE-ResNet50 and SE-ResNext50 to simultaneously predict KL and OARSI grades in knee radiographs. It is computationally heavy due to ensembling. Random forest regression voting approach from the BoneFinder tool[14] is applied to localize knee joints, which is also a traditional machine learning method with lower detection performance. Besides, it uses the whole knee joint images for training models, ignoring the symmetry, semantic information and more detailed features of knee joints that affect the accuracy of grading individual knee OA features. Moreover, these methods all use top1 accuracy to measure prediction performance, ignoring semi-quantitative labels and their ambiguity.

To solve problems above, we propose a SE-ResNext50-32x4d-based Siamese network with adaptive gated feature fusion for individual knee OA features grades prediction. A deep learning method with two cascaded small multi-task networks is presented to localize initial knee joints, which is able to extract more detailed knee features to enhance detection performance compared to traditional machine learning methods. To further reduce redundancy, we propose to further crop detected knee joint regions via an adaptive method to generate main knee joints for follow-up evaluation. Each knee joint is split into the left and right patches equally via its symmetry and fed into a SE-ResNext50-32x4d-based Siamese network with shared weights, extracting more specific and richer features with fewer learning parameters. Adaptive gate mechanism strategy is used to fuse two parts' features before fully connected (FC) layers, which is helpful to capture valuable semantic information and more distinguishable contrast features of two parts. Furthermore, we put forward simultaneous assessment of eight tasks, where the knee OA ($KL \geq 2$)/non-knee OA ($KL \leq 1$) classification task is added, learning more available features for prediction accuracy improvement. One new performance metric that is the top$\pm$1 accuracy is introduced for labels with ambiguity. Specifically, if it is a KL1 knee image and predicted as KL1, KL0 or KL2, the prediction is accepted as accurate. The top$\pm$1 accuracies of OARSI grades are the same. The main contributions in this paper are shown as follows:

(1) Two cascaded small multi-task networks are proposed to localize knee joints for extracting more effective knee features and enhancing detection accuracy.

(2) A SE-ResNext50-32x4d-based Siamese network with shared weights is first used to extract richer features from two knee joints' patches for grading individual knee OA features, making full use of knee joints' symmetry.

(3) An adaptive gated strategy is applied to the feature fusion of the FC layer, capturing more useful semantic information and better contrast features of two patches.

(4) Eight tasks ate simultaneously evaluated, where the knee OA ($KL \geq 2$)/non-knee OA ($KL \leq 1$) classification task is added for the first time to promote feature extraction.

(5) We come up with one new performance metric (i.e., the top$\pm$1 accuracy) for assessing KL and OARSI grades. And our proposed method achieves the state-of-the-art performance in grading individual knee OA features.

## Related works

Several classical studies include knee OA KL-grade diagnosis from X-ray[12,15–19], individual knee OA features assessment from X-ray[20–22], knee OA progression prediction[23] and Magnetic Resonance Imaging (MRI) data analysis[8,9]. As for knee OA KL-grade diagnosis from X-ray, Shamir et al. introduced the WND-CHARM method[15–17], which uses the computer-aided analysis to diagnose early knee OA. Recently, deep learning methods have achieved great success in computer vision fields, such as automatic detection[24], automatic segmentation[25], image recognition[26], video classification[27], image retrieval[28], etc. Compared with traditional approaches, deep learning can directly extract features from data and represent data more effectively. Unsurprisingly, deep learning approaches also revolutionize the field of medical image analysis. Antony et al.[18] used Sobel horizontal image gradient features and SVM to localize knee joint regions. Pre-trained convolutional neural networks, such as VGG16[29], VGG-M-128[30] and CaffeNet[31], via the ImageNet dataset[32] are migrated to perform the fine-tuning on the knee OA KL-grade classification task. However, their knee joints localization method suffer a low detection accuracy. Thus, those knee joint regions cannot be directly used for the subsequent diagnosis task, and manually labeled knee joint regions are utilized. FCN-based method[33] for knee joint localization was introduced by Antony et al.[19], and a six-layer convolutional neural network

with mean square error loss function and the cross-entropy loss function is cascaded to predict knee OA KL-grade. However, it is time-consuming for FCN-based method to generate binary images by segmenting knee regions from each pixel. Later, Tiulpin et al.[12] utilized HOG and SVM method[13] to detect knee joints. Knee joints are divided into symmetric image blocks, which are sent into a 7-layer Siamese convolutional neural network to diagnose knee OA KL-grade. However, these studies merely diagnose knee OA KL-grade, ignoring individual features and the side of knee (lateral or medial). So far, merely a few works have studied individual knee OA features assessment from X-ray. Osteoarthritis Research Society International (OARSI) atlas[34] is the feature-specific approach to grade knee OA severity by grading features, such as FO, TO and JSN. The automatic analysis of individual knee OA features was firstly reported by Oka et al.[20]. Later, Thomson et al.[21] utilized shape and texture descriptors to evaluate the presence of osteophytes and knee OA ($KL \geq 2$). However, the test set they used is relatively small compared to other OA studies. Antony et al.[22] proposed a CNN-based approach for simultaneous analysis of KL and OARSI grades, the prediction accuracy of which needs to be further improved. Tiulpin et al.[14] presented an ensemble model of SE-ResNet50 and SE-ResNext50 to simultaneously assess KL and OARSI grades in knee radiographs, which is time-consuming because of ensembling. They put the whole knee joint into training models without considering knee joints' symmetry. They applies random forest regression voting algorithm[14] for knee detection, which is an traditional machine learning method with lower detection performance. Here, in order to improve knee joint detection accuracy, a deep learning method with two-level cascaded multi-task network is proposed. To extract richer knee joint features and more meaningful semantic information, a SE-ResNext50-32x4d-based Siamese network with shared weights and adaptive gated feature fusion method is proposed to process knee joint patches and simultaneously assess more tasks.

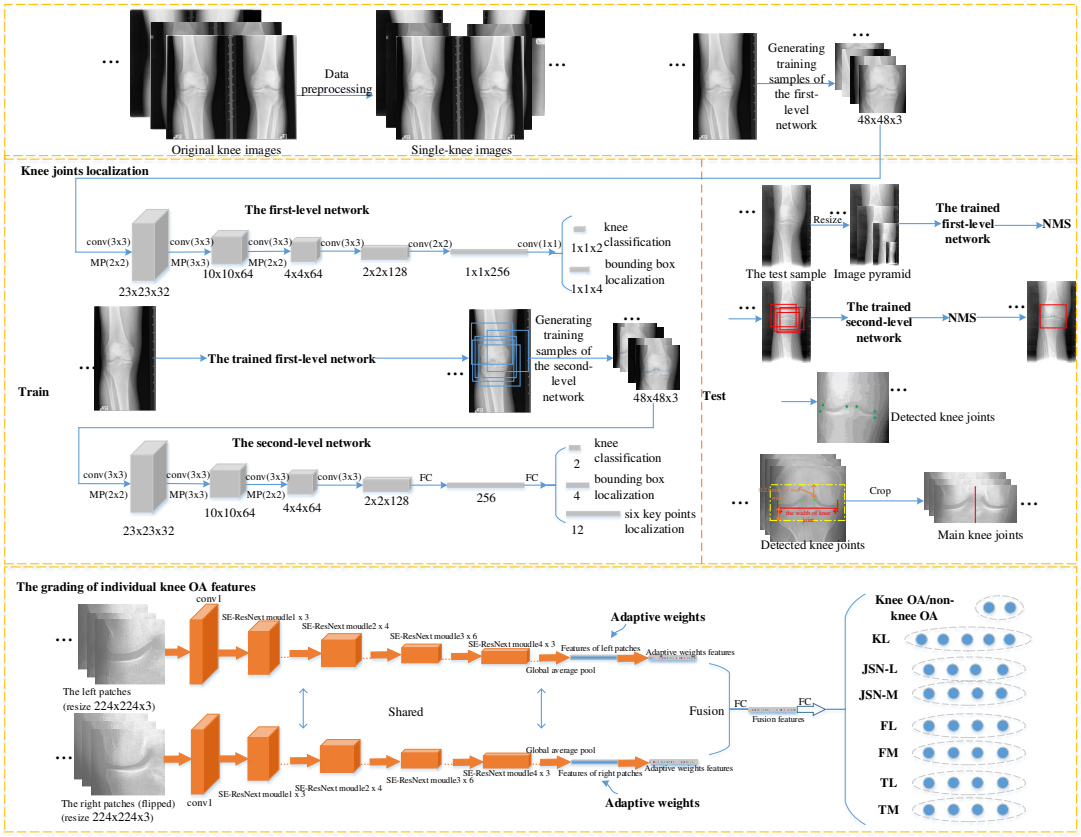



**Figure 2.** The whole process of the proposed method.

## Methods

This study was approved by the Institutional Reviewing Board (IRB) of National Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology, and Xiangya Hospital, Central South University with informed consent obtained from all participants prior to the start of the study. All methods were carried out in accordance with relevant guidelines and regulations. Our experimental data are publicly available, which were approved by the institutional

review board of the University of California San Francisco and obtained the informed consent of all subjects participating in the study. It is described in detail in datasets and data preprocessing part.

The whole process of our proposed method is shown as Figure 2. Original double-knee images are processed and divided into single-knee images. A two-level cascaded multi-task network is trained and tested for localizing knee joints of single-knee images. To further reduce redundancy, located knee joints are cropped again to generate main knee joint regions, each of which is divided into left and right patches with its symmetry. After the right image patch is horizontally flipped, then two patches are fed into a Siamese SE-ResNext50-32x4d Network with shared weights. Adaptive gate mechanism strategy is exploited to fuse contrast features from two patches before fully connected (FC) layers and put them into subsequent FC layers for assessing individual knee OA features grasdes. Eight tasks are simultaneously assessed, where knee OA ($KL \geq 2$)/non-knee OA ($KL \leq 1$) classification task is added for the first time to further enhance feature extraction.

## Datasets and data preprocessing

We utilized two publicly available knee X-ray datasets: the OAI dataset (https://nda.nih.gov/oai/) and the MOST dataset (http://most.ucsf.edu). The OAI dataset we obtained contains the X-ray data from 4796 subjects and their seven follow-up examinations (i.e., baseline, 12-month, 24-month, 36-month, 48-month, 72-month and 96-month). The MOST dataset we obtained includes the X-ray data from 3026 participants and their five follow-up examinations (i.e., baseline, 15-month, 30-month, 60-month and 84-month), which are not belonging to the OAI dataset. Both datasets include double-knee X-rays from men and women aged between 50-79 and 45-79 years old. They were approved by the institutional review board of the University of California San Francisco and the data acquisition sites, which obtained the informed consent of all subjects participating in the study. Because data with missing labels exist in two databases, we select data with KL labels and OARSI scorings for our experiments.

Each double-knee X-ray contains the left and the right knee images in two datasets, which are made up of DICOM format files. Due to the influence of different illumination during shooting, some X-rays have the bright background and dark knees, and some have the opposite. Data preprocessing is first performed to unify knee X-rays, the specific process of which is shown in supplementary Figure 1. First of all, original X-ray knee images with bright background and dark knees are chosen to perform pixel inversion, which means that these X-rays are transformed into images with dark background and bright knees. Therefore, all double-knee images are turned to dark background and bright knees. Then we divide each double-knee image into two single-knee images. Meanwhile, each DICOM format image is converted into the 8-bit uint image. Finally, the histogram equalization is used on all single-knee images. In the end, 24319 and 18634 single-knee images are generated on the OAI dataset and the MOST dataset, respectively.

## Knee joints localization

A two-level cascaded multi-task network is built to localize knee joint regions, which is inspired by MTCNN method[35]. As shown in Figure 2, the knee joints localization network contains two small neural networks. In the training process, training samples are generated first, which are introduced in supplementary information. Training samples of the first-level network are $48 \times 48 \times 3$. Each training sample is sent into the first-level network to sequentially perform three convolutions and maximum pooling operations. Consequently, three convolution operations are connected. In the end, one-dimensional vector about knee/non-knee classification and 4-dimensional bounding box regression vectors of candidate knee joint regions are output. Training samples of the second-level network are also $48 \times 48 \times 3$. The second-level network also performs three convolutions and maximum pooling operations on the $48 \times 48 \times 3$ images. Then one convolution operation and two FC layers are connected behind. Finally, one-dimensional vector about knee/non-knee classification, 4-dimensional bounding box regression vectors of detected knee joints and 12-dimensional vectors of six key points are output. The whole training target is as (1), where $N$ is the number of training samples. $\alpha_{\text{det}}$, $\alpha_{\text{box}}$ and $\alpha_{\text{key points}}$ stand for the importance of knee/non-knee classification task, bounding box regression task and key points localization task, respectively. $\text{Loss}_i^{\text{det}}$, $\text{Loss}_i^{\text{box}}$ and $\text{Loss}_i^{\text{key points}}$ represent the cross-entropy loss of knee/non-knee classification task, the Euclidean loss of bounding box regression task and the Euclidean loss of key point localization task for the $i$-th sample, respectively. In the first-level network, we set $\alpha_{\text{det}} = 1$, $\alpha_{\text{box}} = 0.5$ and $\alpha_{\text{key points}} = 0$. In the second-level network, we set $\alpha_{\text{det}} = 0.8$, $\alpha_{\text{box}} = 0.6$ and $\alpha_{\text{key points}} = 1.5$.

$$\min \sum_{i=1}^{N} \{\alpha_{det} \text{Loss}_i^{\text{det}} + \alpha_{box} \text{Loss}_i^{\text{box}} + \alpha_{keypoints} \text{Loss}_i^{\text{key points}}\}. \tag{1}$$

For the test process, each single-knee image is resized with different scales to generate the image pyramid. The image pyramid is put into the trained first-level network and some candidate knee joint regions are generated. Then highly overlapped candidate regions are merged by the non-maximum suppression (NMS). All reserved candidate regions are fed to the trained second-level network. The second-level network further rejects a few false candidates. Then, NMS is also carried out. Finally, the knee joints and six key points of single-knee images are detected.

## The grading of individual knee OA features

In this subsection, we will specifically describe our SE-ResNext50-32x4d-based Siamese network via adaptive gated feature fusion for the grading of individual knee OA features in plain radiographs (X-ray). Firstly, we perform data processing to further crop detected knee joint areas via six key points and flexibility obtain main knee joint areas for further reducing redundancy. Then, we divide each main knee joint region into left and right patches according to its symmetry and flip the right patch horizontally, which are fed into a SE-ResNext50-32x4d-based Siamese network with shared weights. Finally, adaptive gated feature fusion is used to fuse more distinguishable contrast features of two patches before FC layers.

**Data processing.** Tiulpin et al.[12] selected a fixed position and size area from each located knee joint region for subsequent diagnosis. However, the knee joint width of each person is different. If the knee joint areas are cropped according to the fixed number of pixels, which will lead to inaccurate repositioning of knee joint regions. Relocating initial knee joint areas via the ratio of the knee joint width for each person is able to increase the flexibility and accuracy of knee joints relocation. In this paper, we regard the difference between the maximum and minimum x-coordinates of the six key points as the knee joint width. In the Figure 2, for each detected knee joint from the network of knee joints localization, the maximum and minimum ordinates of the six key points are first found. Then the maximum ordinate increases by 0.2 times of knee joint width as the top of the main knee joint. The minimum ordinate is reduced by 0.2 times of knee joint width as the bottom of the main knee joint. Finally, the main knee joints are obtained with same width and cropped height compared to initially detected knee joint regions.

**The SE-ResNext50-32x4d-based Siamese network.** The effective SE-ResNext50-32x4d-based Siamese network we proposed consists of two SE-ResNext50-32x4d branches. The SE-ResNext50-32x4d branch is built up by a stack of modules, as shown in Figure 2, where a basic SE-ResNext module (see in supplementary Figure 4 (d)) includes the ResNext module and the Squeeze and Excitation (SE)[36] module.

The ResNext block proposed by Xie et al.[37] is a residual block[38] with split-transform-merge strategy in Inception[39,40]. The ResNext block performs a set of transformations, as shown in supplementary Figure 4 (a), where each transformation is set as the bottleneck shaped architecture[37]. Firstly, the vector $x$ is broken up into low-dimensional embeddings before the first $1 \times 1$ layers. Then transformations are performed for low-dimensional embeddings. Finally, all transformations are aggregated. The output of ResNext block can be represented as (2), where $f_i()$ is a function that divides $x$ into a low-dimensional embedding and transforms it. $C$ is defined as cardinality[41], which is the number of aggregated transformations. Here, C is set as 32 and the width of the bottleneck is 4 in SE-ResNext50-32x4d model. As the Figure 4 (b) of supplementary information shown, when the ResNext module uses grouped convolutions[42], it becomes more simple and equivalent to Figure 4 (a) in supplementary information. In the module, 32 $1 \times 1$ layers are replaced by a $1 \times 1$, 128-d layer. Then 32 groups of convolutions are performed in the grouped convolutional layer and finally concatenated as the output.

$$y = x + \sum_{i=1}^{C} f_i(x). \tag{2}$$

The Squeeze and Excitation (SE) unit is introduced by Hu et al.[36] and used to improve the representational capacity of the network by explicitly establishing channel interdependencies of features maps. The SE block is able to selectively highlight value features and suppress less useful ones by feature recalibration, the structure of which is illustrated in supplementary Figure 4 (c). Features P are obtained after one or a series of convolution operations. Then a corresponding SE block follows. Firstly, features P are passed through a squeeze operation, which uses the global average pooling. The feature maps across spatial dimensions are aggregated and channel descriptors are generated as (3), where $c = 1, 2, \ldots C$, $z_c$ represents the $c$-th channel descriptor. Then an excitation operation follows, which aims to acquire dependencies on channels in (4). The first FC layer that is a dimensionality-reduction layer with parameters $W_1$ with reduction ratio $r$ is performed, $W_1 \in R_{C_r \times C}$. A ReLU[43] function is behind and represented as $\tau$. The second FC layer is a dimensionality-increasing layer with parameters $W_2$ and $W_2 \in R_{C \times C_r}$. A sigmoid function as a simple self-gating mechanism is used to produce a corresponding weight for each channel. In the end, the output of the SE block is generated by weighting features P as (5), where $\widetilde{P}_c$ represents the weighted feature of the $c$-th channel.

$$z_c = F_{sq}(p_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} p_c(i, j). \tag{3}$$

$$w = F_{ex}(z, W) = \sigma(W_2 \tau(W_1 z)). \tag{4}$$

$$\widetilde{P}_c = F_{scale}(p_c, w_c) = w_c \cdot p_c. \tag{5}$$

**The adaptive gated feature fusion method.** Inspired by Zhang et al.[44] who proposed a gated multimodal fusion method, we propose an adaptive gated feature fusion method. First of all, we extract features of the left and right knee joint image patches through a SE-ResNext50-32x4d-based Siamese network for selecting useful features and suppressing useless features. We adopt a gate mechanism to adaptively decide weights of the left knee patch features and the right knee patch features for extracting more meaningful comparative features. Then the fused features are obtained as (6).

$$g = \sigma(W_g(f_{left} \bigoplus f_{right})), f = g f_{left} + (1-g) f_{right}. \tag{6}$$

Where $W_g$ is the network parameter. $f_{left}$ and $f_{right}$ stand for the features of left knee patches and right knee patches extracted from the SE-ResNext50-32x4d-based Siamese network, respectively. $\bigoplus$ is the concatenating operation, $\sigma$ is the sigmoid function, $g$ is the weight applied to the $f_{left}$ feature, *(1-g)* is the the weight applied to the $f_{right}$ feature. $f$ is the fused feature of the left and the right knee patches. This is an adaptive learning method, which learns weighted features of each image patch and the correlation between two symmetrical patches. The regions that tasks pay more attention to are learned by the proposed model, which can improve the accuracy of assessing individual knee OA features. Eight tasks are predicted simultaneously, including predictions of KL grades, JSN-L grades, JSN-M grades, FL grades, FM grades, TL grades and TM grades, and knee/non-knee OA classification that we first proposed to add.

### Implementation details

**Knee joint detection model.** Training parameters of the first-level network are set as follows: the *epoch* is 10, the *learning rate (lr)* is 0.001, and the *batch_size* is 500. The training parameters of the second-level network are set as follows: the *epoch* is 10, the *learning rate (lr)* is 0.0001, and the *batch_size* is 500.

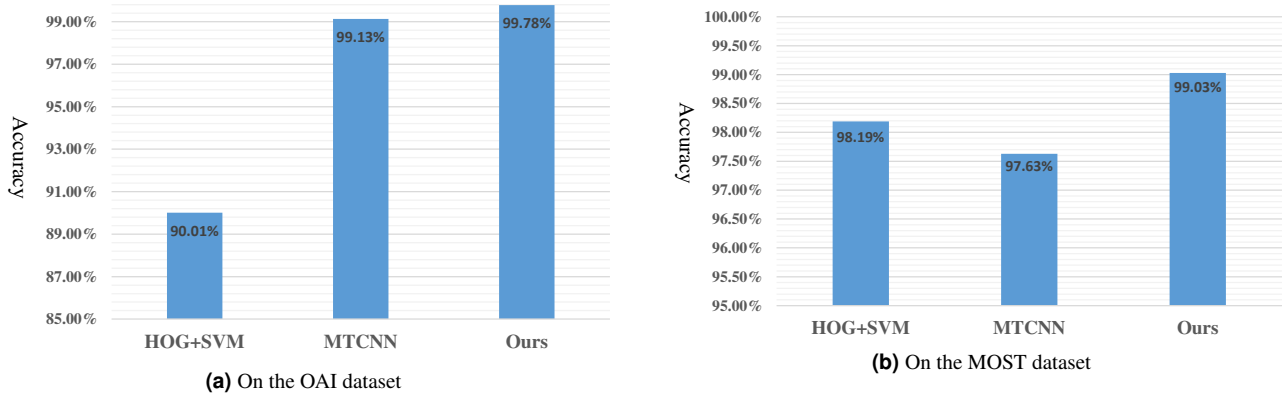
**Grading individual knee OA features model.** We set the training parameters on both datasets as follows[14]: For the first two training epochs, only the FC layers are trained with *learning rate (lr)* of 0.01. Subsequently, the whole network is trained with *lr* of 0.001. *lr* is switched to 0.0001 from the fourth epoch. 20 epochs and Adam optimizer[45] are used in all experiments. To avoid over-fitting problems[46], we use data augmentations[29,38]: illumination contrast enhancement, gamma correction, rotation and translation, etc. Besides, we also use weight decay of 0.0001 and dropout of 0.5 that is inserted before each FC layer.

Our experiments are deployed on the Ubuntu 16.04 platform, replying on Python 3.6, PyTorch 0.4 and 2080Ti GPU.

## Results

### Experimental results and analyses of knee joint detection

The knee joints localization model achieves 99.85% accuracy in validation set, i.e., 1360 out of 1362 are detected. Test results of two datasets are shown in Figure 3. Figure 3 (a) shows that the proposed knee joints localization method is able to detect more knee joints than the HOG+SVM[13] and MTCNN[35] methods on the OAI dataset. The average detection accuracy is 99.78%, which is 0.65% higher than the MTCNN[35] method, and 9.77% higher than the HOG+SVM[13] method. From the Figure 3 (b), the average detection accuracy of our method is 99.03% on the MOST dataset, which is 0.84% higher than that (98.19%) of the HOG+SVM[13] method and 1.4% higher than the MTCNN[35] method. Therefore, our knee joints localization algorithm shows superior performance. Finally, 24265 and 18454 knee joint regions from the OAI dataset and the MOST dataset are detected, respectively, which can be directly used for the subsequent assessment of individual knee OA features.




**(a)** On the OAI dataset

**(b)** On the MOST dataset

**Figure 3.** The detection accuracy comparison of knee joints with different methods on the OAI and MOST datasets.

## Experimental results and analyses of grading individual knee OA features

24265 detected knees from the OAI dataset are divided into 13472 training sets, 2732 validation sets and 8061 test sets. 18454 detected knees from the MOST dataset are divided into 10244 training sets, 2048 validation sets and 6162 test sets. Full description of these data is presented in supplementary Table 1 and Table 2. These detected knee joint regions are relocated to generate critical knee joint areas.

**Comparison with state-of-the-arts.** Table 1 and Table 2 show experimental results of grading individual knee OA features, where ours (Ens.) represents an ensemble model (see supplementary Figure 5) of SE-ResNext50-32x4d and SE-ResNext50-32x4d-based Siamese network for eight tasks. As for the single model on the OAI dataset, the first half of Table 1 shows that our proposed method (ours) is superior to the method proposed by Tiulpin et al.[12], SE-ResNet-50[14] model and SE-ResNext50-32x4d[14] model in most cases for top1 accuracy; we also find that our proposed method (ours) has the better performance than the current best single model (i.e., SE-ResNext50-32x4d) in most cases in terms of top$\pm$1 accuracy from the lower half of Table 1. As for the ensemble model on the OAI dataset, ours (Ens.) is better than ensemble model[14] except for JSN-L and JSN-M in top1 and top$\pm$1 accuracy. Table 1 clarifies that the optimal top1 accuracy of assessing knee OA/non-knee OA, KL-grade, FL-grade, FM-grade, TL-grade, TM-grade, JSN-L-grade and JSN-M-grade are 89.60%, 78.18%, 75.05%, 74.82%, 77.74%, 73.60%, 92.02% and 81.26%, respectively; 97.01%, 95.31%, 94.68%, 97.05%, 97.77%, 98.05% and 98.08% are best top$\pm$1 accuracy of KL-grade, FL-grade, FM-grade, TL-grade, TM-grade, JSN-L-grade and JSN-M-grade on the OAI dataset, most of which are obtained under ours (Ens.). From Table 2, we can observe that ours and ours (Ens.) both outperform previous state-of-the-art models in top1 and top$\pm$1 accuracy in most cases. The highest top1 accuracy of knee OA/non-knee OA, KL-grade, FL-grade, FM-grade, TL-grade, TM-grade, JSN-L-grade and JSN-M-grade are 92.84%, 77.86%, 82.05%, 80.54%, 80.12%, 76.71%, 93.69% and 84.83%, respectively, which are mainly generated by our proposed methods except for JSN-L-grade and JSN-M-grade. Our proposed single model reaches the best top$\pm$1 accuracy of 98.59%, 95.55%, 94.06%, 96.87%, 98.15%, 98.33% in KL-grade, FL-grade, FM-grade, TL-grade, TM-grade, JSN-L-grade, respectively. And the best top$\pm$1 accuracy of 98.31% in JSN-M-grade is achieved under ours (Ens.) method. The single model we proposed even surpasses the ensemble model[14] in most cases on two databases. Thus, we can conclude that our proposed methods have the state-of-the-art performance, which extract more useful, richer features and semantic information, improving top1 and top$\pm$1 accuracy in grading individual knee OA features. In addition, we evaluate our proposed algorithms from Kappa and MSE metrics as shown in supplementary Table 3 and Table 4, where our proposed methods have higher Kappa and lower MSE than existing state-of-the-arts in most cases. Confusion matrices for the KL, OARSI grades prediction and knee OA/non-knee OA binary classification prediction tasks are displayed from Figure 6 to Figure 20 in supplementary information, where predicted results for each grade of each task are shown. Ours and ours (Ens.) methods demonstrate obvious superiority compared with the current best single model[14] and ensemble model[14] for most grades prediction, respectively.

**Table 1.** Performance comparison between the proposed method and other methods on the OAI dataset.

| top1 / Methods | knee OA/non-knee OA | KL | FL | FM | TL | TM | JSN-L | JSN-M |
|---|---|---|---|---|---|---|---|---|
| Tiulpin et al., 2018[12] | - | 63.79% | - | - | - | - | - | - |
| SE-ResNet-50[14] | - | 72.96% | 70.56% | 70.54% | 74.30% | 68.12% | 91.20% | 78.55% |
| SE-ResNext50-32x4d[14] | - | 75.74% | 72.75% | 72.91% | 76.79% | 71.69% | 91.48% | 79.92% |
| Ensemble[14] | - | 76.99% | 73.24% | 73.04% | 76.89% | 71.93% | **92.02%** | **81.26%** |
| Ours | 88.48% | 76.32% | 74.80% | 74.73% | 76.81% | 72.73% | 91.11% | 80.00% |
| Ours (Ens.) | **89.60%** | **78.18%** | **75.05%** | **74.82%** | **77.74%** | **73.60%** | 91.39% | 80.36% |
| top$\pm$1 / Methods | - | KL | FL | FM | TL | TM | JSN-L | JSN-M |
| Tiulpin et al., 2018[12] | - | 88.54% | - | - | - | - | - | - |
| SE-ResNet-50[14] | - | 95.89% | 93.44% | 92.74% | 95.89% | 97.26% | 97.84% | 97.98% |
| SE-ResNext50-32x4d[14] | - | 96.22% | 94.52% | 94.02% | 96.07% | 97.43% | 98.02% | 97.90% |
| Ensemble[14] | - | 96.35% | 94.42% | 94.05% | 96.17% | 97.72% | **98.05%** | **98.08%** |
| Ours | - | 96.63% | **95.31%** | 94.59% | 96.91% | 97.67% | 97.89% | 97.75% |
| Ours (Ens.) | - | **97.01%** | 95.12% | **94.68%** | **97.05%** | **97.77%** | 97.85% | 97.85% |

**Ablation study.** SE-ResNext50-32x4d model represents one seven task model of the entire image as input. However, Siamese SE-ResNext50-32x4d model stands for a seven task model of two image blocks as input. As shown in Table 3, Siamese SE-ResNext50-32x4d method exceeds SE-ResNext50-32x4d method in most cases in terms of top1 and top$\pm$1 accuracy on the OAI dartaset. Table 4 displays that Siamese SE-ResNext50-32x4d method has higher top1 accuracy than SE-ResNext50-32x4d method in KL-grade, FL-grade, FM-grade, TL-grade and TM-grade; and the better top$\pm$1 accuracy is acquired with Siamese SE-ResNext50-32x4d method in KL-grade, TL-grade and JSN-M-grade compared with SE-ResNext50-32x4d method. Thus, it has been proved that two image blocks as input is beneficial to extract more sufficient features than the whole knee joint as input, enhancing the classification performance.

In order to verify the effectiveness of adding knee OA/non-knee OA classification task, on the one hand, we compare

**Table 2.** Performance comparison between the proposed method and other methods on the MOST dataset.

| top1 / Methods | knee OA/non-knee OA | KL | FL | FM | TL | TM | JSN-L | JSN-M |
|---|---|---|---|---|---|---|---|---|
| Tiulpin et al., 2018[12] | - | 68.73% | - | - | - | - | - | - |
| SE-ResNet-50[14] | - | 74.39% | 79.32% | 78.33% | 76.87% | 71.19% | 93.20% | 84.00% |
| SE-ResNext50-32x4d[14] | - | 75.22% | 78.98% | 78.72% | 78.04% | 73.11% | 93.43% | 84.16% |
| Ensemble[14] | - | 76.13% | 80.31% | 79.23% | 78.42% | 73.56% | **93.69%** | **84.83%** |
| Ours | 92.58% | 76.94% | **82.05%** | **80.54%** | **80.12%** | 76.32% | 93.54% | 83.77% |
| Ours (Ens.) | **92.84%** | **77.86%** | 80.88% | 79.91% | 79.52% | **76.71%** | 93.31% | 84.19% |
| top±1 / Methods | - | KL | FL | FM | TL | TM | JSN-L | JSN-M |
| Tiulpin et al., 2018[12] | - | 90.28% | - | - | - | - | - | - |
| SE-ResNet-50[14] | - | 98.26% | 93.72% | 92.16% | 94.55% | 97.73% | 98.17% | 98.18% |
| SE-ResNext50-32x4d[14] | - | 98.20% | 95.16% | 93.28% | 96.45% | 97.84% | 98.26% | 98.26% |
| Ensemble[14] | - | 98.43% | 94.64% | 93.20% | 95.80% | 97.97% | 98.30% | 98.28% |
| Ours | - | **98.59%** | **95.55%** | **94.06%** | **96.87%** | **98.15%** | **98.33%** | 98.21% |
| Ours (Ens.) | - | 98.12% | 94.86% | 93.72% | 96.59% | 98.10% | 98.20% | **98.31%** |

SE-ResNext50-32x4d to SE-ResNext50-32x4d with 8 tasks, whose inputs both are the whole relocated knee joint image. On the other hand, Siamese SE-ResNext50-32x4d with 8 tasks is compared with Siamese SE-ResNext50-32x4d, whose inputs are the left and the right patches of relocated knee joints. From Table 3 we can find that SE-ResNext50-32x4d with 8 tasks is superior to SE-ResNext50-32x4d for top1 accuracy in most cases on the OAI dataset. Siamese SE-ResNext50-32x4d with 8 tasks method has better performance than Siamese SE-ResNext50-32x4d in terms of the top1 accuracy of KL-grade, FL-grade, TM-grade, JSN-L-grade and JSN-M-grade and except for the top±1 accuracy of TL-grade. As for the MOST dataset, Table 4 demonstrates that SE-ResNext50-32x4d with 8 tasks outperforms SE-ResNext50-32x4d except for the JSN-L-grade's top1 accuracy. But the top±1 accuracy of SE-ResNext50-32x4d with 8 tasks method is less than that of SE-ResNext50-32x4d. Siamese SE-ResNext50-32x4d with 8 tasks method beats Siamese SE-ResNext50-32x4d in terms of the top1 accuracy of FL-grade, FM-grade, TL-grade and TM-grade and in terms of the top±1 accuracy of FL-grade, FM-grade, TM-grade and JSN-L-grade. In general, adding knee OA/non-knee OA binary classification in the final prediction layer is helpful to improve prediction performance of individual knee OA features grading by facilitating extracting richer knee OA features.

In Table 3, the top1 accuracy of assessing individual knee OA features via our proposed method that uses adaptively gated feature fusion method is superior to that of Siamese SE-ResNext50-32x4d with 8 tasks on the OAI dataset under most circumstances. Meanwhile, Table 4 describes that our method has better performance in all top1 accuracy compared with Siamese SE-ResNext50-32x4d with 8 tasks. The top±1 accuracy of our method is higher than that of Siamese SE-ResNext50-32x4d with 8 tasks apart from for FM-grade. Therefore, adaptively gated feature fusion strategy further boosts effects of individual knee OA features assessment, which highlights more valuable contrast features and semantic information of key regions.

**Table 3.** Performance Comparison of ablation experiments on the OAI dataset.

| top1 / Methods | knee OA/non-knee OA | KL | FL | FM | TL | TM | JSN-L | JSN-M |
|---|---|---|---|---|---|---|---|---|
| SE-ResNext50-32x4d[14] | - | 75.74% | 72.75% | 72.91% | 76.79% | 71.69% | 91.48% | 79.92% |
| SE-ResNext50-32x4d with 8 tasks | **89.53%** | **77.19%** | 72.83% | 72.73% | 76.34% | 71.18% | **91.53%** | 79.94% |
| Siamese SE-ResNext50-32x4d | - | 73.66% | 74.42% | 74.71% | **77.48%** | 72.11% | 89.83% | 76.79% |
| Siamese SE-ResNext50-32x4d with 8 tasks | 89.11% | 75.97% | 74.69% | 74.51% | 76.80% | **73.08%** | 91.12% | 78.71% |
| Ours | 88.48% | 76.32% | **74.80%** | **74.73%** | 76.81% | 72.73% | 91.11% | **80.00%** |
| top±1 / Methods | knee OA/non-knee OA | KL | FL | FM | TL | TM | JSN-L | JSN-M |
| SE-ResNext50-32x4d[14] | - | 96.22% | 94.52% | 94.02% | 96.07% | 97.43% | **98.02%** | **97.90%** |
| SE-ResNext50-32x4d with 8 tasks | - | **97.15%** | 93.92% | 93.85% | 96.70% | 97.37% | 97.93% | 97.69% |
| Siamese SE-ResNext50-32x4d | - | 96.79% | 95.11% | 94.02% | **97.39%** | 97.74% | 97.52% | 97.32% |
| Siamese SE-ResNext50-32x4d with 8 tasks | - | 97.12% | **95.66%** | **94.86%** | 97.15% | **97.99%** | 97.93% | 97.75% |
| Ours | - | 96.63% | 95.31% | 94.59% | 96.91% | 97.67% | 97.89% | 97.75% |

## Discussion

In this study, we propose a SE-ResNext50-32x4d-based Siamese network with adaptive gated feature fusion strategy to assess individual knee OA features grades. Based on results above, we conclude the following observation and discussions:

(1) Two cascaded small multi-task networks are designed to locate knee joints, the average detection accuracy of which achieves 99.78% on the OAI dataset and 99.03% on the MOST dataset. Our detection model has capability of extracting more discriminative features compared with previous methods, enjoying obvious superiority. Thus, the proposed detection approach

**Table 4.** Performance Comparison of ablation experiments on the MOST dataset.

| top1 / Methods | knee OA/non-knee OA | KL | FL | FM | TL | TM | JSN-L | JSN-M |
|---|---|---|---|---|---|---|---|---|
| SE-ResNext50-32x4d[14] | - | 75.22% | 78.98% | 78.72% | 78.04% | 73.11% | 93.43% | 84.16% |
| SE-ResNext50-32x4d with 8 tasks | 92.19% | 77.00% | 79.60% | 78.92% | 78.38% | 76.00% | 93.20% | **84.70**% |
| Siamese SE-ResNext50-32x4d | - | **77.04**% | 79.93% | 79.03% | 79.07% | 75.49% | 93.07% | 83.84% |
| Siamese SE-ResNext50-32x4d with 8 tasks | 92.53% | 76.76% | 80.69% | 79.96% | **80.12**% | 76.19% | 92.83% | 83.22% |
| Ours | **92.58**% | 76.94% | **82.05**% | **80.54**% | **80.12**% | **76.32**% | **93.54**% | 83.77% |
| top±1 / Methods | - | KL | FL | FM | TL | TM | JSN-L | JSN-M |
| SE-ResNext50-32x4d[14] | - | 98.20% | 95.16% | 93.28% | 96.45% | 97.84% | 98.26% | 98.26% |
| SE-ResNext50-32x4d with 8 tasks | - | 97.96% | 94.40% | 93.27% | 95.88% | 97.57% | 97.83% | 98.26% |
| Siamese SE-ResNext50-32x4d | - | **98.64**% | 95.07% | 92.71% | 96.54% | 97.52% | 97.99% | **98.33**% |
| Siamese SE-ResNext50-32x4d with 8 tasks | - | 97.89% | 95.47% | **94.22**% | 96.51% | 97.63% | 98.25% | 97.97% |
| Ours | - | 98.59% | **95.55**% | 94.06% | **96.87**% | **98.15**% | **98.33**% | 98.21% |

is an effective tool for locating knee joints, which are applicable for subsequent diagnosis.

(2) As well known, the features of knee OA are concentrated around the knee joint spaces. The original located knee joints may contain some regions that contribute nothing. To reduce redundancy, we further crop located knee joints via six key points to generate critical knee joints.

(3) A SE-ResNext50-32x4d-based Siamese network is first to be used to grade individual knee OA features, where ResNext module and SE module help to capture more useful features from knee images. Furthermore, located whole knee joints are divided into two patches according to their symmetry, which are fed into our Siamese network with shared weights to extract more detailed features.

(4) An adaptive gated strategy is applied to the feature fusion layer to further suppress useless information and highlight valuable information, which helps to capture richer semantic information and obtain better contrast features of two patches.

(5) In order to fully extract knee joints features, we add the knee OA(KL$\geq$2)/non-OA(KL$\leq$1) binary classification task to the other seven tasks we predicted simultaneously. The binary classification of knee OA/non-OA not only enhances other tasks' prediction performance but also is vital for doctors' preliminary clinical diagnosis.

(6) We introduce a new performance metric that is top$\pm$1 accuracy to assess KL and OARSI grades. Due that the development of knee OA is successive and expert evaluation is subjective, KL and OARSI ground-truth labels with discrete values have ambiguity. Therefore, we propose one new standard that the prediction belonging to true label's adjacent grades is also regarded as an accurate one. To verify our methods, we compare our proposed methods of grading individual knee OA features grades with existing methods, our proposed method achieves promising results under different evaluation metrics.

Here, we merely consider the KL, OARSI grades prediction (i.e., JSN-L, JSN-M, FL, FM, TL and TM) and knee OA/non-OA binary classification tasks with sufficient data. Some additional OARSI features are not considered at all, such as medial tibial attrition, medial tibial sclerosis, lateral femoral sclerosis, etc. In the future, more OARSI features could be studied to provide additional clinical advice. Currently, knee joints detection and grading prediction are separate steps. Future work will focus on investigating an end-to-end deep learning system by combining these steps. Moreover, Magnetic Resonance Imaging (MRI) images will be used in the feature, which contains more information. In conclusion, this study demonstrates the automatic grading of individual knee OA features. Despite it has some shortcomings, we believe that the proposed approach has potential to become a useful tool in clinical OA trials and provide better quantitative information for doctors.

## References


1. Yoo, T. K., Kim, D. W., Choi, S. B. & Park, J. S. Simple scoring system and artificial neural network for knee osteoarthritis risk prediction: a cross-sectional study. *PloS one* **11**, e0148724 (2016).
2. Oka, H. *et al.* Fully automatic quantification of knee osteoarthritis severity on plain radiographs. *Osteoarthr. Cartil.* **16**, 1300–1306 (2008).
3. Shamir, L. *et al.* Early detection of radiographic knee osteoarthritis using computer-aided analysis. *Osteoarthr. Cartil.* **17**, 1307–1312 (2009).
4. Shamir, L. *et al.* Knee x-ray image analysis method for automated detection of osteoarthritis. *IEEE Transactions on Biomed. Eng.* **56**, 407–415 (2008).
5. Arden, N. & Nevitt, M. C. Osteoarthritis: epidemiology. *Best practice & research Clin. rheumatology* **20**, 3–25 (2006).
6. Puig-Junoy, J. & Zamora, A. R. Socio-economic costs of osteoarthritis: a systematic review of cost-of-illness studies. In *Seminars in arthritis and rheumatism*, vol. 44, 531–541 (Elsevier, 2015).

7. Braun, H. J. & Gold, G. E. Diagnosis of osteoarthritis: imaging. *Bone* **51**, 278–288 (2012).

8. Pedoia, V. *et al.* 3d convolutional neural networks for detection and severity staging of meniscus and pfj cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. *J. Magn. Reson. Imaging* **49**, 400–410 (2019).

9. Norman, B., Pedoia, V. & Majumdar, S. Use of 2d u-net convolutional neural networks for automated cartilage and meniscus segmentation of knee mr imaging data to determine relaxometry and morphometry. *Radiology* **288**, 177–185 (2018).

10. Kellgren, J. & Lawrence, J. Radiological assessment of osteo-arthrosis. *Annals rheumatic diseases* **16**, 494 (1957).

11. Altman, R. D., Hochberg, M., Murphy, J. W., Wolfe, F. & Lequesne, M. Atlas of individual radiographic features in osteoarthritis. *Osteoarthr. cartilage* **3**, 3–70 (1995).

12. Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P. & Saarakkala, S. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. *Sci. reports* **8**, 1727 (2018).

13. Tiulpin, A., Thevenot, J., Rahtu, E. & Saarakkala, S. A novel method for automatic localization of joint area on knee plain radiographs. In *Scandinavian Conference on Image Analysis*, 290–301 (Springer, 2017).

14. Tiulpin, A. & Saarakkala, S. Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks. *Diagnostics* **10**, 932 (2020).

15. Shamir, L. *et al.* Wndchrm–an open source utility for biological image analysis. *Source code for biology medicine* **3**, 13 (2008).

16. Orlov, N. *et al.* Wnd-charm: Multi-purpose image classification using compound image transforms. *Pattern recognition letters* **29**, 1684–1693 (2008).

17. Shamir, L. *et al.* Wnd-charm: Multi-purpose image classifier. *Astrophys. Source Code Libr.* (2013).

18. Antony, J., McGuinness, K., O'Connor, N. E. & Moran, K. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, 1195–1200 (IEEE, 2016).

19. Antony, J., McGuinness, K., Moran, K. & O'Connor, N. E. Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks. In *International conference on machine learning and data mining in pattern recognition*, 376–390 (Springer, 2017).

20. Oka, H. *et al.* Normal and threshold values of radiographic parameters for knee osteoarthritis using a computer-assisted measuring system (koacad): the road study. *J. Orthop. Sci.* **15**, 781–789 (2010).

21. Thomson, J., O'Neill, T., Felson, D. & Cootes, T. Detecting osteophytes in radiographs of the knee to diagnose osteoarthritis. In *International Workshop on Machine Learning in Medical Imaging*, 45–52 (Springer, 2016).

22. Antony, A. J. *Automatic quantification of radiographic knee osteoarthritis severity and associated diagnostic features using deep convolutional neural networks*. Ph.D. thesis, Dublin City University (2018).

23. Tiulpin, A. *et al.* Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci. Reports* **9**, 1–11 (2019).

24. Nguyen, C. C., Tran, G. S., Nghiem, T. P., Burie, J.-C. & Luong, C. M. Real-time smile detection using deep learning. *J. Comput. Sci. Cybern.* **35**, 135–145 (2019).

25. Liu, C. *et al.* Automatic segmentation of the prostate on ct images using deep neural networks (dnn). *Int. J. Radiat. Oncol. Biol. Phys.* **104**, 924–932 (2019).

26. Kong, F. Facial expression recognition method based on deep convolutional neural network combined with improved lbp features. *Pers. Ubiquitous Comput.* 1–9 (2019).

27. Tran, D., Wang, H., Torresani, L. & Feiszli, M. Video classification with channel-separated convolutional networks. *arXiv preprint arXiv:1904.02811* (2019).

28. Wiggers, K. L., Britto Jr, A. S., Heutte, L., Koerich, A. L. & Oliveira, L. S. Image retrieval and pattern spotting using siamese neural network. *arXiv preprint arXiv:1906.09513* (2019).

29. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

**30.** Chatfield, K., Simonyan, K., Vedaldi, A. & Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531* (2014).

**31.** Jia, Y. *et al.* Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, 675–678 (ACM, 2014).

**32.** Russakovsky, O. *et al.* Imagenet large scale visual recognition challenge. *Int. journal computer vision* **115**, 211–252 (2015).

**33.** Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440 (2015).

**34.** Altman, R. D. & Gold, G. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthr. cartilage* **15**, A1–A56 (2007).

**35.** Zhang, K., Zhang, Z., Li, Z. & Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **23**, 1499–1503 (2016).

**36.** Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141 (2018).

**37.** Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1492–1500 (2017).

**38.** He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).

**39.** Szegedy, C. *et al.* Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9 (2015).

**40.** Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).

**41.** Cantor, G. Ueber unendliche, lineare punktmannichfaltigkeiten. *Math. Annalen* **21**, 51–58 (1984).

**42.** Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105 (2012).

**43.** Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814 (2010).

**44.** Zhang, Q., Fu, J., Liu, X. & Huang, X. Adaptive co-attention network for named entity recognition in tweets. In *Thirty-Second AAAI Conference on Artificial Intelligence* (2018).

**45.** Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *Comput. Sci.* (2014).

**46.** Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal machine learning research* **15**, 1929–1958 (2014).


## Acknowledgements


This work has been supported by the National Key Research and Development Program of China (Grant No. 2018YFB0204301). We would also like to thank the OAI database (https://nda.nih.gov/oai/) sponsored by the National Institutes of Health (part of the Department of Health and Human Services) and the MOST database (http://most.ucsf.edu) sponsored by the National Institutes of Health / National Institute on Aging (part of the Department of Health & Human Services), respectively.


## Author contributions statement

Kang Wang originated the idea of the study. Kang Wang, Xin Niu and Yong Dou designed the study. Kang Wang performed the experiments and wrote the manuscript. Xin Niu and Yong Dou provided the technical feedback. Dongxing Xie and Tuo Yang provided the clinical feedback. All authors participated in the manuscript writing and editing.

## Additional information

**Competing interests statement**
The authors declare no competing interests.
**Non-financial competing statement**
The authors declare non-financial competing interests.

# Figures

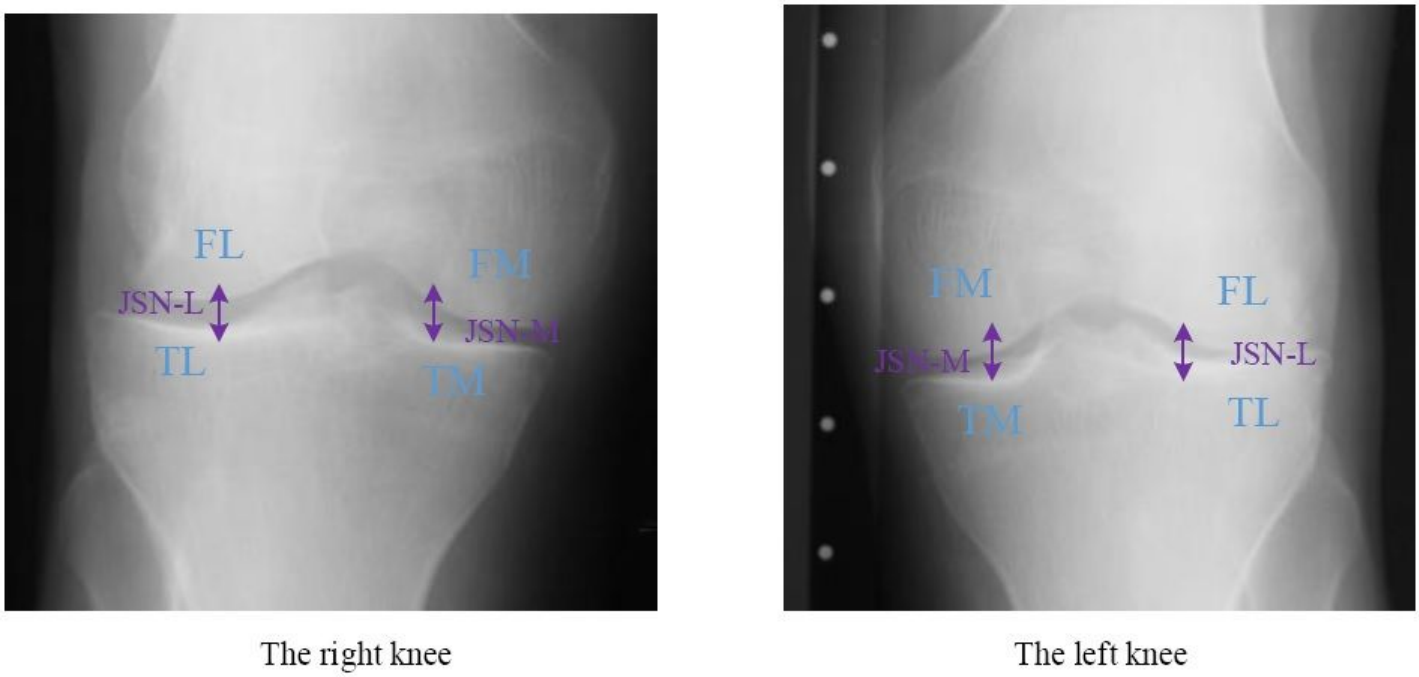
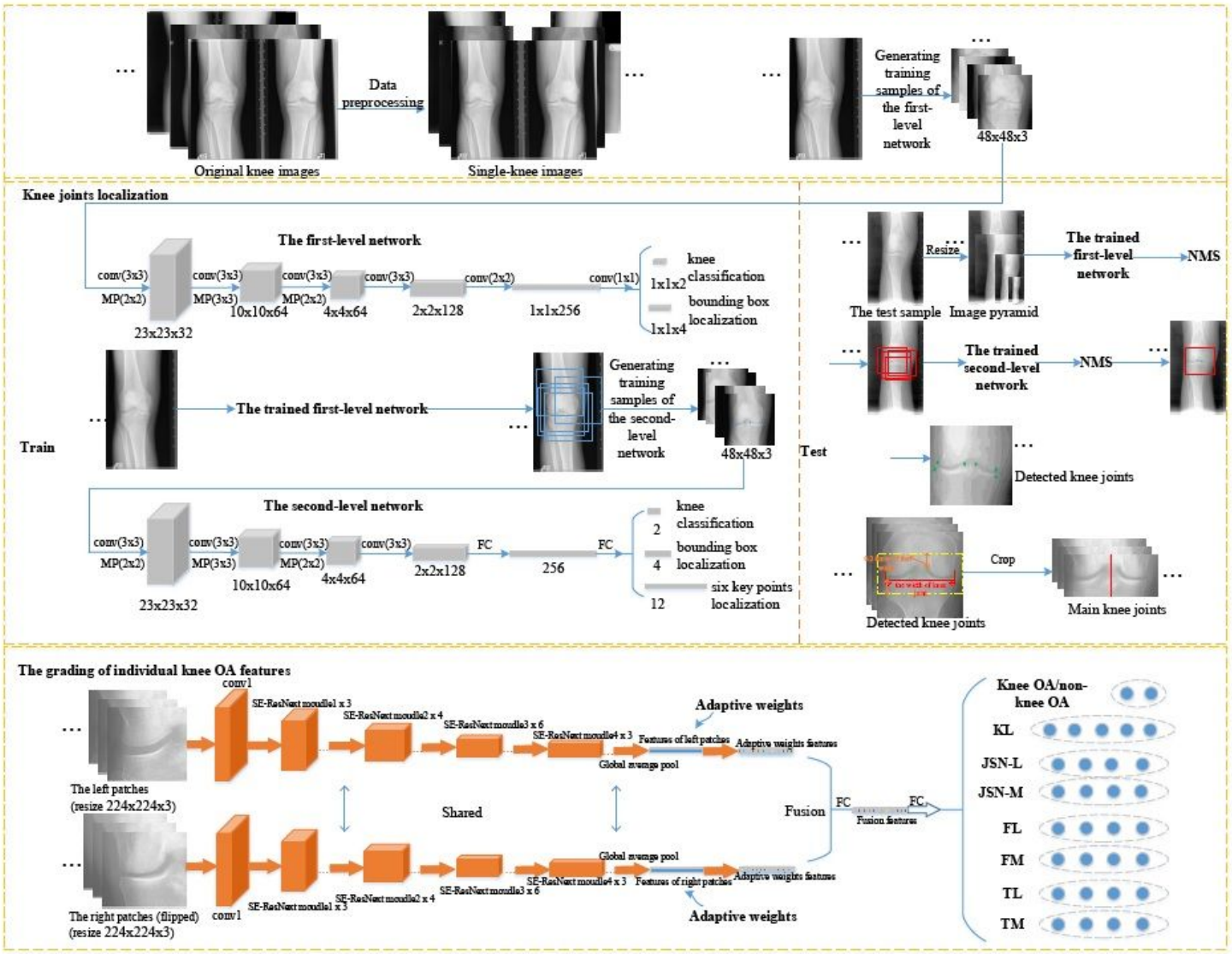


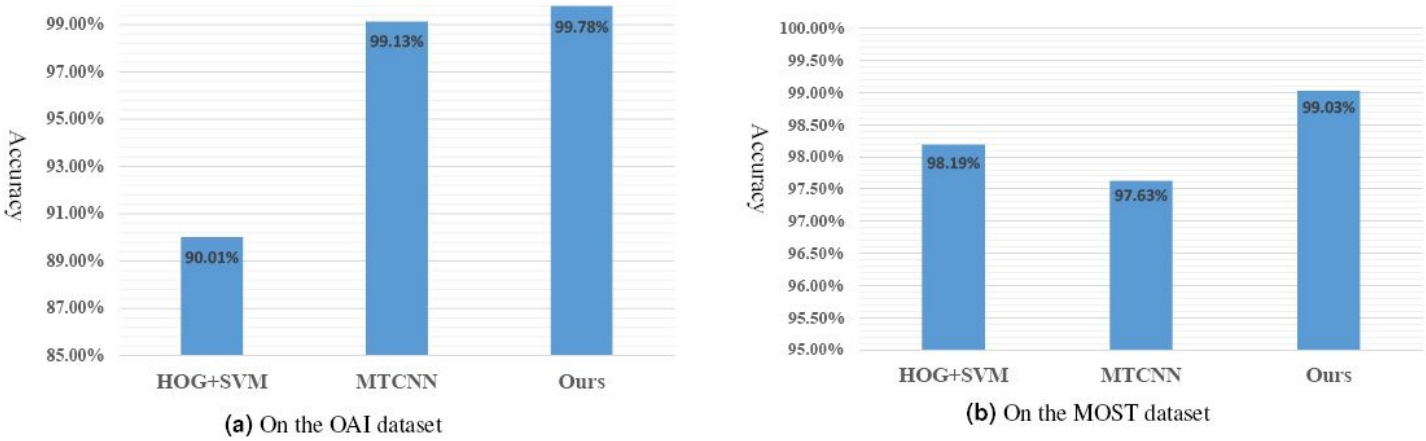

**Figure 1**

The specific features in knee images.




## Figure 2

The whole process of the proposed method.




(a) On the OAI dataset

(b) On the MOST dataset

## Figure 3

The detection accuracy comparison of knee joints with different methods on the OAI and MOST datasets.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.pdf