Published in final edited form as:
Magn Reson Med. 2019 May ; 81(5): 3330–3345. doi:10.1002/mrm.27627.

SUSAN: Segment Unannotated image Structure using Adversarial Network

Fang Liu, Ph.D.[1,2]

[1]Department of Radiology, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Madison, Wisconsin 53705–2275

Abstract

Purpose: To describe and evaluate a segmentation method using joint adversarial and segmentation convolutional neural network (CNN) to achieve accurate segmentation using unannotated magnetic resonance (MR) image datasets.

Methods: A segmentation pipeline was built using joint adversarial and segmentation network. A CNN technique called cycle-consistent generative adversarial network (CycleGAN) was applied as the core of the method to perform unpaired image-to-image translation between different MR image datasets. A joint segmentation network was incorporated into the adversarial network to obtain additional functionality for semantic segmentation. The fully-automated segmentation method termed as SUSAN was tested for segmenting bone and cartilage on two clinical knee MR image datasets using images and annotated segmentation masks from an online publicly available knee MR image dataset. The segmentation results were compared using quantitative segmentation metrics with the results from a supervised U-Net segmentation method and two registration methods. The Wilcoxon signed-rank test was used to evaluate the value difference of quantitative metrics between different methods.

Results: The proposed method SUSAN provided high segmentation accuracy with results comparable to the supervised U-Net segmentation method (most quantitative metrics having $p>0.05$) and significantly better than a multi-atlas registration method (all quantitative metrics having $p<0.001$) and a direct registration method (all quantitative metrics having $p<0.0001$) for the clinical knee image datasets. SUSAN also demonstrated the applicability for segmenting knee MR images with different tissue contrasts.

Conclusion: SUSAN performed rapid and accurate tissue segmentation for multiple MR image datasets without the need for sequence specific segmentation annotation. The joint adversarial and segmentation network and training strategy have promising potential applications in medical image segmentation.

Keywords

Deep Learning; Adversarial Network; Segmentation; MRI; Image Annotation

[2]Corresponding Author: Fang Liu, Ph.D., Department of Radiology, Wisconsin Institutes for Medical Research, 1111 Highland Avenue, Madison, Wisconsin 53705-2275, Phone: 612-222-9728, FAX: 608-263-5112, leoliuf@gmail.com.

INTRODUCTION

Segmentation of magnetic resonance (MR) images is a fundamental step in many medical imaging based applications. Traditionally, image segmentation is performed by having experienced users scroll through stacks of two-dimensional (2D) images and manually segmenting regions-of-interest (ROIs) among adjacent tissues. However, manual segmentation is time-consuming and is influenced by the level of human expertise and errors due to distraction and fatigue associated with human interpretation(1–4). Therefore, manual segmentation is subject to inter- and intra- observer variability which likely leads to inconsistent segmentation results(3,4). There has been much recent interest in developing semi- and fully-automated techniques for segmenting MR images(5). The majority of recently proposed methods for fully-automated segmentation have utilized model-based and atlas-based approaches(5,6). Although these methods have shown promising results, both approaches rely on a priori knowledge of object shapes and thus might perform poorly in situations in which there is high subject variability and significant differences of local features. In addition, these methods require high computation cost which results in relatively long segmentation times.

Recent implementation of deep convolutional neural networks (CNNs) in image processing has been shown to have significant impacts on medical image segmentation (7). Deep CNN-based methods have achieved state-of-the-art performance in many medical image segmentation tasks including segmenting brain tumors(8,9), tissues (10,11), and multiple sclerosis lesions (12), cardiac (13,14), liver(15), and lung(16) tissues, and musculoskeletal tissues such as bone and cartilage(17–19). On the other hand, medical image segmentation is typically seen as a multi-class labeling problem which is closely related to the supervised semantic segmentation described in most segmentation CNN studies. In particular, convolutional encoder-decoder (CED) networks have proven to be highly efficient in the medical image domain. This type of network typically consists of a paired encoder and decoder where the encoder performs image compression and feature extraction and the decoder reconstructs pixel-wise classification labels using encoder outputs. Ronneberger et al. (20) developed U-Net which has a 2D CED structure with skip connections between the encoder and decoder. The U-Net transfers feature maps from the encoder to the decoder and concentrates them to obtain up-sampled feature maps through deconvolution. U-Net was first proposed for segmenting neuronal structures in electron microscopy and was later adapted for many medical image segmentation tasks. Badrinarayanan et al. (21) proposed a 2D CED called SegNet which is built upon the deep structure of the VGG16 network (22) and features a unique up-sampling approach in the decoder using pre-stored max-pooling indices from the encoder. This network offers an efficient alternative to deconvolution for recovering high resolution image features and achieved top performance in multiple segmentation challenges. Expanding on the capabilities of 2D CEDs, such as U-Net and SegNet, recently proposed three-dimensional (3D) CEDs extend convolutional kernels into the slice dimension for volumetric image data and attempt to incorporate full spatial information for improved segmentation performance(23,24). Other segmentation networks using multi-scale and multi-patch based structures have also been proposed and have proven to be quite useful for segmenting 3D image datasets(9,25,26). More recently, further improvement for CNN-based image segmentation was achieved by using adversarial training, where a dedicated CNN network was introduced to correct generated segmentation maps from the ground truth(27). A few pilot studies demonstrated great performance using adversarial training for segmenting brain lesions(28), structures(29), and prostate cancer(30) on MR images, chest organs(31) on X-ray images, and breast cancer on histopathology images(32).

Network training of segmentation CNNs typically requires images and paired annotation data representing pixel-wise tissue labels referred to as masks. The pixel-wise correlation between image pixels and tissue masks is used for supervised training of the segmentation CNNs for learning useful image features. However, the supervised training of highly efficient CNNs with deeper structure and more network parameters requires a large amount of training images and paired tissue masks. Moreover, the creation of tissue masks typically requires individuals with medical expertise to annotate a large number of training image datasets which could be extremely expensive and time consuming(33). Although a trained segmentation CNN may perform well for one type of MR sequence, the applicability of the CNN for segmenting the same tissues on images acquired using other MR sequences is typically poor. Therefore, it is necessary to retrain the CNN using new annotation data specific to each MR sequence. Thus, there is great need to develop a generalized CNN-based segmentation method which would be applicable for a wide variety of MR image datasets with different tissue contrasts.

The purpose of our study was to develop and evaluate a generalized CNN-based method for fully-automated segmentation of different MR image datasets using a single set of annotated training data. A technique called cycle-consistent generative adversarial network (CycleGAN) (34) is applied as the core of the proposed method to perform image-to-image translation between MR image datasets with different tissue contrasts. A segmentation network is incorporated into the adversarial network to obtain additional segmentation functionality. We termed the proposed method as SUSAN standing for Segmenting Unannotated image Structure using Adversarial Network and evaluated SUSAN for segmenting bone and cartilage on two clinical knee MR image datasets acquired at our institution using only a single set of annotated data from a publicly available knee MR image dataset.

THEORY

Adversarial Network for Image-to-Image Translation

Our work is in line with the method of CycleGAN which was recently proposed for unpaired image-to-image translation for natural images (34). In CycleGAN, images from two image domains can be translated to exchange image contrasts, features and patterns (e.g. translate horse into zebra, apple into orange and vice versa) using two key techniques including Cycle Consistency and Generative Adversarial Newark (GAN). Our work is closely related to the basic framework of CycleGAN. The main concept of SUSAN is to translate the reference images which have high quality segmentation annotation into the target images which have no segmentation annotation. Our hypothesis is that given successful translation from the reference image contrast to the target image contrast, the annotation data used to train the reference images can be applied to train the target images using supervised learning.

Transitivity is a useful tool for many computer science applications including human language translation(35,36), dense semantic alignment(37), image co-segmentation(38) and 3D shape matching(39). In the CycleGAN setup (Figure 1a), the idea of cycle consistency is to apply transitivity to force forward-backward data cycle consistency between two image domains. Mathematically, the minimization of *cycle consistency loss* is to learn forward (F) and backward (B) CNN mapping functions

$$\begin{aligned} F &: X \rightarrow Y \\ B &: Y \rightarrow X \end{aligned} \qquad [1]$$

between two image domains X and Y, so that if an image is translated from one domain to the other and back again, the image should look identical to the original image. Therefore, for forward cycle consistency, there should be $x \rightarrow F(x) \rightarrow B(F(x)) \approx x$ and for backward cycle consistency, there should be $y \rightarrow B(y) \rightarrow F(B(y)) \approx y$. The cycle consistency loss term can be formulated as

$$L_{cyc}(F, B) = \mathbf{E}_{x\sim P(x)}[\|B(F(x)) - x\|_N] + \mathbf{E}_{y\, -\, P(y)}[\|F(B(y)) - y\|_N] \qquad [2]$$

where $\mathbf{E}_{i\sim P(i)}[\cdot]$ is the expectation of a function and the data distribution of image i in domain I is denoted as $i \sim P(i)$. The $\|\cdot\|_N$ is typically chosen to be l_1 norm for image-to-image translation (34).

GAN is another key concept for CycleGAN. Many recent GAN studies have achieved impressive results in a variety of image subspecialties including image inpainting (40), text to image synthesis(41), and image generation(42,43). The idea of GAN is to use an *adversarial loss* to force CNN mapping function generating synthetic images that are indistinguishable from the real images. In the current study, adversarial loss can be incorporated into the CNN training as a joint loss term. For example, for the CNN mapping function $F: X \rightarrow Y$, a multiple layer CNN discriminator $D_Y(y)$ is defined to identify real versus synthetic images. Mathematically, this discriminator outputs a scalar representing the probability that y comes from the real domain Y rather than the forward mapping output $F(x)$. The adversarial loss term is thus formatted as

$$L_{gan}^{f}(F, D_Y) = \mathbf{E}_{y\,\sim\,P(y)}[\log D_Y(y)] + \mathbf{E}_{x\,\sim\,P(x)}[\log(1 - D_Y(F(x)))] \qquad [3]$$

In addition, the CycleGAN also introduced a CNN discriminator $D_X(x)$ for the backward mapping with adversarial loss as:

$$L_{gan}^{b}(B, D_X) = \mathbf{E}_{x\,\sim\,P(x)}[\log D_X(x)] + \mathbf{E}_{y\,\sim\,P(y)}[\log(1 - D_X(B(y)))] \qquad [4]$$

The full objective function for CycleGAN is given as:

$$L(F, B, D_X, D_Y) = \lambda_{\sigma x} L_{\infty}(F, B) + \lambda_{gan}\left(L_{gan}^{f}(F, D_Y) + L_{gan}^{b}(B, D_X)\right) \quad [5]$$

where λ_{cyc} and λ_{gan} are weight factors for the cycle consistency loss and adversarial loss term, respectively, to balance the data fidelity of image translation and the GAN quality. The full objective function is trained in a two-player minimax game, namely F and B aim to minimize this loss function against the adversary D_X and D_Y that try to maximize it as

$$\widetilde{F}, \widetilde{B} = \underset{F, B}{\operatorname{argmin}} \underset{D_X, D_Y}{\max} L(F, B, D_X, D_Y) \quad [6]$$

In other words, F and B try to generate synthetic images that look similar to real images, while discriminator D_X and D_Y try to distinguish synthetic images from real images. In theory, successful training of this network can result in mapping CNNs capable of generating synthetic images indistinguishable from real images in the original domain(34,44). Namely, the translated images in each image domain just look like real images from that domain.

Joint Segmentation Network

To make use of image-to-image translation, SUSAN incorporates additional segmentation networks into the CycleGAN structure for jointly training image translation and segmentation (Figure 1b). Our hypothesis is that the image translation can be augmented by adding supervised segmentation information and that jointly training image translation and segmentation can produce improved results. To formulate the problem, the CNN mapping function is modified to create a dual-output network. Namely, the mapping function is designed to not only output translated images but also full size tissue segmentation masks. Therefore, Eq. [1] can be rewritten as

$$\begin{aligned} F&: X \rightarrow \left(Y, M_f\right) \\ B&: Y \rightarrow \left(X, M_b\right) \end{aligned} \quad [7]$$

where M_f and M_b are output segmentation masks for the forward and backward CNN, respectively. It is assumed that all the images in image domain X have high quality segmentation masks M_x which can be used for supervised training. The segmentation loss is incorporated into the full objective function as

$$L_{seg}(F, B) = \mathbf{E}_{x \sim P(x)}[\ell(F(x), M_x)] + \mathbf{E}_{x \sim P(x)}[\ell(B(F(x)), M_x)] \quad [8]$$

where $\ell(\cdot)$ is a loss metric for evaluating pixel-wise similarity between the output mask and the ground truth mask, e.g. multi-class cross entropy (45). According to Eq. [8], supervised segmentation training is performed for all image x in domain X and all translated synthetic image $F(x)$ which have origins from image domain X. The full objective function incorporating segmentation loss can be extended from Eq. [5] into

$$L(F, B, D_X, D_Y) = \lambda_{cyc} L_{cyc}(F, B) + \lambda_{seg} L_{seg}(F, B) + \lambda_{gan}\left(L_{gan}^{f}(F, D_Y) + L_{gan}^{b}(B, D_X)\right) \quad [9]$$

where λ_{seg} is a weight factor for the segmentation term. Note that once the joint training process is complete, for image domain X, the final segmentation CNN is simply given by the forward CNN as

$$M_f = F(x), x \in X \quad [10]$$

Likewise, for image domain Y which has no explicit annotated segmentation mask, the final segmentation result is given by the backward CNN output as

$$M_b = B(y), y \in Y \quad [11]$$

METHODS

Network Implementation

A U-Net architecture (20) was adapted from a GAN-based image-to-image translation study (46) for performing the CNN mapping functions (i.e. F and B) between the two image domains. This U-Net structure is composed of an encoder network and a decoder network. The encoder is used to achieve efficient data compression while probing robust and spatial invariant image features. A decoder network with a mirrored structure of the encoder is applied following the encoder network output for restoring desirable image features. Multiple symmetric shortcut connections are added to transfer features from the encoder to the decoder to enhance mapping performance. Such a CNN structure has shown impressive results for image translation and segmentation in many recent studies(20,46,47). In the current study, the U-Net structure is modified to enable dual outputs, and this new design is referred to as R-Net. Namely, the network is bifurcated following the last up-sampling layer in the decoder. One segmentation branch of R-Net uses a multi-class soft-max classification layer as the final layer (48) which produces class probabilities for each individual pixel at the same image resolution as the input image. The image translation branch of R-Net uses a convolution layer as the last layer to generate gray scale image also matching the input image resolution. An illustration of the R-Net is shown in Figure 2a.

Similar to the original CycleGAN work, the network architecture developed in PatchGAN (46) is used for discriminator networks (i.e. D_X and D_Y), which aim to classify whether overlapping image patches are real or fake in the adversarial process. An illustration of PatchGAN is shown in Figure 2b. Such a patch-level discriminator architecture has the advantage of fewer parameters allowing for more efficient network training and has been shown in many recent GAN studies to have impressive results for differentiating real versus synthetic images(46,49,50).

Image Datasets

Three knee MR image datasets were used to evaluate SUSAN. The reference image dataset consisted of the online knee MR images available from The Segmentation of Knee Images 2010 (SKI10, www.ski10.org) workshop competition hosted by the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference in 2010 (51). The image dataset in SKI10 consisted of 60 sagittal fat-suppressed T1-weighted spoiled gradient-echo (T1-SPGR) knee images, all of which had high quality multi-class tissue masks manually generated by segmentation experts with the following values: 0=background, 1=femur, 2=femoral cartilage, 3=tibia, and 4=tibial cartilage.

Two clinical knee MR image datasets acquired at our institution were used as target images. The study was performed in compliance with Health Insurance Portability and Accountability Act (HIPAA) regulations, with approval from our Institutional Review Board, and with a waiver of written informed consent. Sagittal knee image datasets acquired with two MR sequences were retrospectively obtained on patients undergoing a clinical knee MR examination at our institution using a 3T scanner (Discovery MR750, GE Healthcare) and 8-channel phased-array extremity coil. The first image dataset consisted of a sagittal fat-suppressed T2weighted fast spin-echo (T2-FSE) sequence acquired on 60 patients. The second image dataset consisted of a sagittal proton density-weighted fast spin-echo (PD-FSE) sequence acquired on 60 patients. The imaging parameters for both MR sequences are summarized in Table 1. Manual bone and cartilage segmentation of the clinical knee images was performed by a musculoskeletal research scientist with 8 years of segmentation experience. Note that the manual segmentation for the clinical image datasets was used only for ground truth comparison and was not included in the network training for SUSAN. Multi-class tissue masks were created for the clinical knee image datasets using the same labelling values as used for the SKI10 image dataset.

Network Training

All the input 2D images from the clinical knee MR image datasets were first cropped to enclose as much of the knee joint as possible while removing excessive image background, leading to an approximate 512 × 400 image matrix size. Due to the GPU global memory limit, all the images were further resampled to 256 × 256 matrix size using bilinear interpolation before they were sent to the network for training and evaluation. Image normalization was also performed for each subject by subtracting the mean value of the entire image and then normalizing by the standard deviation of the image signal intensity. When training the network, the network weights were initialized using the initialization scheme of He et al. (52) and updated using Adam algorithm (53) with a fixed learning rate of 0.0002 and trained in a mini-batch manner with three image slices in a single mini-batch. Multi-class cross entropy loss was applied for the segmentation branch (in Eq. [8]) and the l_1 loss for the image translation branch (in Eq. [2]) in the R-Net. During training iteration, a two-step training strategy was applied where CNN mapping functions (i.e. F and B) and adversarial discriminators (i.e. D_X and D_Y) were updated separately in an alternating manner. A default set of parameters for the weight factors in the full objective function was empirically selected and included $\lambda_{cyc} = 10$, $\lambda_{gan} =1$ and $\lambda_{seg} = 5$ in the Eq. [9] for the clinical knee image datasets. To investigate the influence of different weights for the segmentation branch on the performance of image translation and subsequent segmentation, the experiments were also performed for $\lambda_{seg} = 0.5$ and 10, respectively, while maintaining other parameters the same.

For the network training of SUSAN, a split of 50/10 subjects (5506/1083 slices) in the SKI10 image dataset and a split of 35/5 randomly selected subjects in the PD-FSE (1343/201 slices) and T2-FSE (937/146 slices) image datasets were used for training and validation, respectively, while the remaining 20 subjects in the PD-FSE and T2-FSE image datasets were used for hold-out evaluation. A total iteration steps corresponding to 20 epochs for the SKI10 image dataset were carried out for training convergence. The best model was selected in which the calculated loss was the lowest for the set of validation images. To compare SUSAN with a standard supervised method, direct U-Net training for tissue segmentation was performed using the clinical knee image datasets, and the corresponding multi-class tissue masks created using manual segmentation. The supervised learning was performed by removing the image translation branch from the RNet while keeping the aforementioned training procedure the same. To compare SUSAN with state-of-the-art conventional segmentation methods, a multi-atlas registration algorithm from the Knee Segmentation and Registration Toolkit (KSRT, https://bitbucket.org/marcniethammer/ksrt)(54) was evaluated. Multiple atlases were built from all 60 subjects in SKI10 image dataset and the registration workflow and parameters were kept the same as the default setting implemented in the source code and stated in the original paper(54). In addition, to compare SUSAN with direct registration approach between different image contrasts, a registration algorithm using Elastix software and registration parameters (http://elastix.isi.uu.nl/wiki.php) as described in (55) was also evaluated. The segmentation of bone and cartilage was obtained for the clinical datasets by directly registering the images into each individual SKI10 subject image as one template. The segmentation accuracy of SUSAN, the supervised U-Net method, the multi-atlas registration and the direct registration method (value reported using the best segmentation from all templates in SKI10 datasets) was evaluated on the 20 hold-out test subjects in the PD-FSE and T2-FSE image datasets.

SUSAN was implemented in Python language (v2.7, Python Software Foundation, Wilmington, Del). The CNNs were designed using the Keras package (56) running Tensorflow computing backend (57) at a 64-bit Ubuntu Linux system. All training and evaluation were performed on a computer server with an Intel Xeon W3520 quad-core CPU, 32 GB DDR3 RAM, and one Nvidia GeForce GTX 1080Ti graphic card with total 3584 CUDA cores, and 11GB GDDR5 RAM.

Evaluation of Synthetic Images

To evaluate the quality of synthetic images, the FCN-score, a quantitative metric used in the original CycleGAN paper (34) was implemented for image assessment. While direct evaluation of the generative model is challenging, the idea is to use a pre-trained semantic segmentation classifier to measure the discriminability of synthetic images against the ground truth images. The assumption is that the segmentation classifier trained on the ground truth images should be able to segment synthetic images at a high accuracy if the synthetic images are realistic(34,46). More specifically, a FCN-8s network (45) was adapted for the segmentation classifier, and was trained on the real PD-FSE and T2-FSE dataset, respectively, using the same aforementioned training procedure. The trained model was then applied to segment 20 synthetic PD-FSE and T2-FSE images and 20 real hold-out clinical knee images for comparison, respectively. The Per-pixel accuracy was reported as FCN-score with the definition as follows

$$ACC = \frac{\sum_i n_{ii}}{\sum_i t_i} \qquad [12]$$

where n_{ii} is the number of pixels of class i correctly predicted to belong to class i, and t_i is the total number of pixels of class i. The FCN-score implementation in this study was adapted from the online code at https://github.com/phillipi/pix2pix/tree/master/scripts/eval_cityscapes.

Evaluation of Segmentation Accuracy

Quantitative metrics were used to evaluate the accuracy of the segmentation methods on the different clinical knee MR image datasets. To evaluate the volumetric segmentation accuracy, the Dice coefficient (DC) was used for bone and cartilage and was defined as

$$\mathrm{DC} = \frac{2|S \cap R|}{|S| + |R|} \qquad [13]$$

where S and R represented the CNN segmentation and the manual segmentation ground truth, respectively. The Dice coefficient ranges between 0 and 1 with a value of 1 indicating a perfect segmentation and a value of 0 indicating no overlap at all. The volumetric overlap error (VOE) was also calculated to evaluate the accuracy of cartilage segmentation. The VOE was defined as

$$\mathrm{VOE} = 1 - \frac{|S \cap R|}{|S \cup R|} \qquad [14]$$

with a smaller VOE value indicating a more accurate segmentation. The VOE values were calculated within a ROI drawn in each of three consecutive central slices on the medial and lateral tibial plateau, and medial and lateral femoral condyles. To evaluate the surface overlap between the segmented masks and the ground truth, the Average Symmetric Surface Distance (ASSD) was calculated for bone and cartilage. The ASSD was defined as

$$\mathrm{ASSD} = \frac{\sum_{s \,\in\, \partial(S)\mathrm{min}} \min_{r \,\in\, \partial(R)} \left\| s - r \right\| + \sum_{r \,\in\, \partial(R)} \min_{s \,\in\, \partial(S)} \left\| r - s \right\|}{|\partial(S)| + |\partial(R)|} \qquad [15]$$

where $\partial(\cdot)$ means the boundary of the segmentation set. A small ASSD value typically indicates similar surface boundaries and great surface overlap, thereby reflecting a more accurate segmentation. For statistical analysis, a paired non-parametric Wilcoxon signed-rank test was used to compare the DC, VOE and ASSD values between SUSAN, the supervised U-Net method, and two registration methods at a pre-defined significance level of $p<0.05$.

RESULTS

The overall training time required for SUSAN was approximate 6.7 hours for each clinical knee MR image dataset given the computing hardware in the current study. However, once the training was complete, fully-automated segmentation was rapid with a mean computing time of 0.2 min for all image slices in the PD-FSE and T2-FSE image datasets. In contrast, the multi-atlas registration method took a mean computing time of 5.2 hours for all image slices in the PD-FSE and T2-FSE image datasets. The direct registration method took less time with a mean computing time of 0.5 hours for all image slices in the PD-FSE and T2-FSE image datasets using one template.

Figure 3 and 4 show an example of unpaired image-to-image translation for converting the SKI10 T1-SPGR image contrast to the PD-FSE and T2-FSE image contrasts (denoted as synthetic PD-FSE and T2-FSE) and vice versa. Note that in Figure 3 and 4, the subject was randomly selected from the individual image datasets for demonstration. The T1-SPGR images had low fat signal in bone and subcutaneous soft tissue due to fat suppression and bright cartilage and muscle signal. The PD-FSE images had high fat signal in bone and subcutaneous soft tissue and low cartilage and muscle signal. The T2-FSE images had low signal for all tissues except synovial fluid which was bright. Despite the dramatic differences in tissue contrasts between the three image datasets, SUSAN was capable of translating the varying MR contrasts with a good visual appearance as a result of the incorporated CycleGAN functionality. The qualitative observation was supported by the quantitative evaluation. Figure 5 further demonstrated an example of synthetic PD-FSE and T2-FSE images from a T1-SPGR image at different epochs with different FCN-scores calculated using ACC metric. In addition, there were FCN-scores (mean ± standard deviation) of 0.73 ± 0.03 and 0.79 ± 0.05 for all the synthetic PD-FSE and T2-FSE images, and 0.78 ± 0.02 and 0.83 ± 0.04 for all the real hold-out PDFSE and T2-FSE images, respectively, indicating a high similarity between the real images and the synthetic images from SUSAN. The influence of different weights of the segmentation branch on the image translation and segmentation accuracy are shown in Table 2. Although a large weight (λ_{seg} =10) on the segmentation branch was not significantly different than the default moderate weight (λ_{seg} =5) for translating and segmenting the PD-FSE and T2-FSE images, it was evident that a substantially reduced weight (λ_{seg} =0.5) for the segmentation branch performed noticeably worse for the image translation and segmentation on both the PD-FSE and T2-FSE images.

The averaged values (mean ± standard deviation) of DC, VOE and ASSD are shown in Table 3 for bone and cartilage segmentation for the 20 hold-out test subjects in the PD-FSE and T2-FSE image datasets. Although the segmentation accuracy was significantly higher for the supervised U-Net method than SUSAN for femoral cartilage (DC: p=0.008, VOE: p=0.008) on PD-FSE images and for tibia bone (DC: p=0.008, ASSD: p=0.002) on T2-FSE images. SUSAN provided overall comparable segmentation performance to the supervised U-Net method for the PD-FSE and T2-FSE image datasets while requiring no sequence specific annotated training data. The multi-atlas registration method performed significantly worse (p<0.001 for all bone and cartilage) than both SUSAN and the supervised U-Net method. The direct registration method also performed significantly worse (p<0.0001 for all bone and cartilage) than deep learning methods.

Figure 6 shows examples of bone and cartilage segmentation performed on sagittal PD-FSE images of the knee in a 56 year old male subject with mild knee osteoarthritis The segmentation results from SUSAN demonstrated good agreement with the overall contours of the ground truth. There was also good agreement of the overall shape between SUSAN and supervised U-Net method. For SUSAN, there was a bone segmentation accuracy of DC 0.94 and 0.95 for the femur and tibia respectively indicating small deviations from the ground truth. There was a cartilage segmentation accuracy of ASSD 0.84mm and 0.70mm for femoral and tibial cartilage respectively indicating good cartilage segmentation. For the supervised U-Net method, there was a bone segmentation accuracy of DC 0.94 and 0.96 for the femur and tibia respectively; there was a cartilage segmentation accuracy of ASSD 0.75mm and 0.65mm for femoral and tibial cartilage respectively. The deep learning methods outperformed the multi-atlas registration method. Although the multi-atlas method provided a similar bone shape to the ground truth for femur and tibia, the cartilage segmentation using the atlases built from SKI10 images failed on the PD-FSE images. For the direct registration, the segmentation result for both bone and cartilage was even worse than the multi-atlas registration and thus not shown in this figure.

Figure 7 shows examples of segmentation performed on sagittal T2-FSE images of the knee in a 64 year old male subject with knee osteoarthritis. Tissue segmentation was challenging for this subject due to the thin articular cartilage on the femur and tibia. The segmentation results from SUSAN demonstrated good agreement with the overall contours of the ground truth although a few small clusters of misclassification in bone and cartilage (white arrows) were observed due to ambiguous tissue contrast and low signal-to-noise ratio. For SUSAN, there was a bone segmentation accuracy of DC 0.96 and 0.93 for the femur and tibia respectively indicating small differences from the ground truth. There was a cartilage segmentation accuracy of ASSD 0.59mm and 0.71mm for femoral and tibial cartilage respectively indicating good cartilage segmentation. For the supervised U-Net method, there was a bone segmentation accuracy of DC 0.96 and 0.96 for the femur and tibia respectively; there was a cartilage segmentation accuracy of ASSD 0.46mm and 0.54mm for femoral and tibial cartilage respectively. The deep learning methods also substantially outperformed the multi-atlas registration method for bone and cartilage segmentation on T2-FSE images. The direct registration method failed to provide reasonable segmentation for both bone and cartilage for this subject thus not shown in this figure.

DISCUSSION

Our study described a novel adversarial CNN-based segmentation method that provided rapid and accurate segmentation for multiple MR image datasets using only a single set of annotated training data. This technique eliminates the need to retrain the segmentation CNNs using new annotation data specific to each MR sequence. The approach integrates the basic functions of the CycleGAN technique for image-to-image translation and utilizes an additional semantic segmentation network for joint image translation and segmentation. The segmentation results for segmenting two clinical knee MR image datasets suggested that this proposed method SUSAN utilizing joint adversarial and segmentation network can achieve high accuracy with performance comparable to a state-of-the-art supervised CNN method and better than registration methods. SUSAN was also highly time efficient with an average segmentation time less than one minute which is much shorter than the segmentation times of currently used fully-automated atlas-based and model-based segmentation methods.

A limitation of CNN-based segmentation methods is the need for adequate training data which may often be difficult to acquire. Many recent deep learning studies have demonstrated continuously increasing segmentation performance with increasing training dataset size(58,59). Since large training datasets are difficult to collect, several methods have been proposed to alleviate the problem. One approach is to implement data augmentation techniques where the size of the training data could be efficiently increased by applying random geometrical transformation to the original image data such as rotation, translation and deformation. These data augmentation techniques have become useful techniques in many recent deep learning medical image studies to overcome small dataset size problem and to reduce the training overfitting(8,60). Transfer learning and fine-tuning techniques are other popular solutions to small training dataset size problem. Instead of initializing the network with random weights during training, the pre-trained weight values from the same image domain, different medical image domains, or even natural image datasets can be used to initiate the training process. Then, the training can be conducted in a supervised manner for the entire network or a few network layers using new annotated training data specific to the segmentation task. Many recent studies have shown that fine-tuning can improve performance and reduce requirement for training data and this performance improvement increases with reducing training dataset size(61,62). However, in both data augmentation and transfer learning techniques, annotated training data for the specific image dataset is still required, regardless of its size. We proposed a new approach to alleviate the problem by utilizing image translation to create domain specific training data. Our results suggest that a joint adversarial and segmentation network can be used for rapid and accurate segmentation for multiple MR image datasets using only a single set of annotated training data. Our method provides an alternative approach to perform medical image segmentation in circumstances in which collecting annotated training data is challenging, expensive, or not possible.

Although SUSAN was only evaluated for image translation for knee MR image datasets, there is no inherent limitation preventing this approach from translating images for other anatomic structures and for other imaging modalities. For example, it would be useful to translate knee joint cartilage segmentation into cartilage segmentation of the hip joint since annotating hip joint cartilage is extremely difficult due to the thin cartilage and closely opposing articular surfaces. One concern for such image translation is the significant geometric difference between image domains. Although the applied U-Net is widely accepted for mapping image contrast and texture as a results of efficient convolutional encoding and decoding design (46,47), successfully translation among different anatomy might still require further design of the generator architecture tailored for handling both contrast translation and geometric transformation. In addition, translating image contrasts among different imaging modalities would be useful in applications such as positron emission tomography (PET)/MR attenuation correction, where there is the need to generate synthetic computed tomography (CT) images from MR images for photon attenuation calculation (63,64). In current study, the use of adversarial training and cycle consistency regularization as key techniques in the CycleGAN framework was proven to be effective to learn mutually correlated image features with unpaired data in medical image domain. The adversarial learning ensures the translated images falling into the same data distribution of the target images; the cycle consistency prevents the degeneracy of the adversarial process from generating hallucinated image features (34). This cross domain adaptation in CycleGAN imposes both shared-latent space constraints and information consistency to encourage CNNs to learn mutually correlated image features in different data distribution. Although current method achieves compelling results for translating MR images between different contrasts, the interpretation of the translated image contrast requires careful attentions in clinical practice. It should be noted that a typical assumption about high similarity of data characteristics in high dimensional feature space between training datasets might fail when pathology occurs in one contrast but not the other. Therefore, the synthetic images cannot be used as reliable diagnostic replacements for real images. Although the synthetic images in SUSAN provide sufficient information for image segmentation, they may not reflect true tissue contrasts for pathological conditions. Comprehensive assessment for the effectiveness and applicability of MR image translation using large prospective image datasets is needed for various diseases and tissue structures. In addition, SUSAN uses entirely unpaired image data thus might suffer from model collapse in adversarial training (34). Future comparison between current unpaired translation and the paired image-to-image translation which has stronger constraints in supervised learning is of great interests. Given a fraction of annotated data at minimal cost, a certain form of weakly supervised image-to-image translation incorporating both paired and unpaired information might provide further performance improvement.

Current SUSAN used 2D CNNs which can potentially limit image features within a single slice and may cause segmentation bias when contrast is inconsistent across slices in an image volume. This was the likely causes of the tissue misclassification noted in our results for the clinical knee MR images in Figure 7 where tissue contrast was contaminated by the large noise level. Incorporating 3D information could be helpful for image translation to a target image with ambiguous tissue contrast. Multi-planner CNN methods, usually referred to as 2.5D methods, and fully 3D CNN methods have been shown in many studies to improve segmentation accuracy and efficiency (9,23,24). However, high dimensional CNN methods typically require extensive computing resources, such as abundant GPU memory, which is a prohibitively limiting factor in many studies. In our network configuration, the requirement for hardware resources when running high dimensional CNNs was even more severe since we implemented both forward and backward CNNs, which doubled the total network size. Alternatively, implementing suitable post-processing methods to adjust the segmentation results from a 2D CNN output is also applicable to take into account the 3D contextual relationships of the full image volume. Recent studies have demonstrated the successful use of fully-connected 3D conditional random field (CRF) to regularize segmentation boundaries at tissue interfaces (9,11,18). The 3D surface shape-based morphological deformable approach has also proven to be highly efficient to maintain desirable geometrical shape for segmented objects in combination with CNN-based segmentation method (17,18). Since these post-processing steps typical require no GPU computation, they could be very efficient and require little computational costs.

Our study has several limitations. First, the current study adapted the U-Net and PatchGAN structures and did not compare other CNN structures. Newly developed CNNs (65,66) will be explored in future studies. Second, the network training parameters were selected based on heuristic information. Tuning the weighting factors in the objective function is important to control the balance between image translation and structure segmentation, and is likely dependent of specific studies. Although the results from this study demonstrated that a moderately weighted segmentation branch improved the image translation (Table 2), comprehensive parameter optimization would be necessary in future studies to investigate further improvement for the performance. Third, future studies investigating the influence of imaging parameters on the image translation and segmentation accuracy are also needed. Finally, the current method did not compare other methods which use adversarial networks for image segmentation.

CONCLUSIONS

In conclusion, our study described a new fully-automated CNN-based segmentation method which integrates joint adversarial and segmentation CNNs to segment MR images with different tissue contrasts using a single set of annotated training data. Our method was shown to provide rapid and accurate segmentation of bone and cartilage for clinical knee MR image datasets comparable to a state-of-the-art supervised CNN method. Additional studies are needed to evaluate potential applications of SUSAN for other anatomical structures and for other imaging modalities. The new technique may further improve the applicability and efficiency of CNN-based segmentation of medical images while eliminating the need for large amounts of annotated training data.

ACKNOWLEDGEMENT

The authors thank the nurses and research technologists who supported this work at the University of Wisconsin-Madison. The authors thank Dr. Richard Kijowski at the University of Wisconsin-Madison who helped language editing the manuscript in this work.

REFERENCES

1. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116–1128. [PubMed: 16545965]
2. McWalter EJ, Wirth W, Siebert M, von Eisenhart-Rothe RMO, Hudelmaier M, Wilson DR, Eckstein F. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. Osteoarthr. Cartil. [Internet] 2005;13:48–53. doi: 10.1016/j.joca.2004.09.008.
3. Ashton EA, Takahashi C, Berg MJ, Goodman A, Totterman S, Ekholm S. Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. J. Magn. Reson. Imaging [Internet] 2003;17:300–308. doi: 10.1002/jmri.10258.
4. Shim H, Chang S, Tao C, Wang J-H, Kwoh CK, Bae KT. Knee Cartilage: Efficient and Reproducible Segmentation on High-Spatial-Resolution MR Images with the Semiautomated Graph-Cut Algorithm Method. Radiology [Internet] 2009;251:548–556. doi: 10.1148/radiol.2512081332.
5. Sharma N, Aggarwal LM. Automated medical image segmentation techniques. J. Med. Phys. [Internet] 2010;35:3–14. doi: 10.4103/0971-6203.58777.

6. Pedoia V, Majumdar S, Link TM. Segmentation of joint and musculoskeletal tissue in the study of arthritis. Magn. Reson. Mater. Physics, Biol. Med. [Internet] 2016;29:207–221. doi: 10.1007/s10334-0160532-9.

7. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med. Image Anal. 2017;42:60–88. doi: 10.1016/j.media.2017.07.005. [PubMed: 28778026]

8. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation using Convolutional Neural Networks in MRI Images. IEEE Trans Med Imaging [Internet] 2016. doi: 10.1109/TMI.2016.2538465.

9. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 2017;36:61–78. doi: 10.1016/j.media.2016.10.004. [PubMed: 27865153]

10. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Isgum I. Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging [Internet] 2016. doi: 10.1109/TMI.2016.2548501.

11. Zhao G, Liu F, Oler JA, Meyerand ME, Kalin NH, Birn RM. Bayesian convolutional neural network based MRI brain extraction on nonhuman primates. Neuroimage [Internet] 2018;175:32–44. doi: 10.1016/j.neuroimage.2018.03.065.

12. Brosch T, Tang L, Yoo Y, Li D, Traboulsee A, Tam R. Deep 3D Convolutional Encoder Networks with Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. IEEE Trans Med Imaging [Internet] 2016. doi: 10.1109/TMI.2016.2528821.

13. Avendi MR, Kheradvar A, Jafarkhani H. Fully automatic segmentation of heart chambers in cardiac MRI using deep learning. J. Cardiovasc. Magn. Reson. [Internet] 2016;18:2–4. doi: 10.1186/1532-429X18-S1-P351.

14. Avendi MR, Kheradvar A, Jafarkhani H. Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. Magn. Reson. Med. [Internet] 2017;78:2439–2448. doi: 10.1002/mrm.26631.

15. Ferdinand Christ P, Ezzeldin Elshaer MA, Ettlinger F, et al. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. Proc. 19th Int. Conf. Med. Image Comput. Comput. Assist. Interv. [Internet] 2016:1–8. doi: 10.1007/978–3-31946723–8.

16. Zha W, Kruger SJ, Johnson KM, Cadman RV., Bell LC, Liu F, Hahn AD, Evans MD, Nagle SK, Fain SB. Pulmonary ventilation imaging in asthma and cystic fibrosis using oxygen-enhanced 3D radial ultrashort echo time MRI. J. Magn. Reson. Imaging [Internet] 2017. doi: 10.1002/jmri.25877.

17. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging. Magn. Reson. Med. [Internet] 2017:DOI: 10.1002/mrm.26841. doi: 10.1002/mrm.26841.

18. Zhou Z, Zhao G, Kijowski R, Liu F. Deep Convolutional Neural Network for Segmentation of Knee Joint Anatomy. Magn. Reson. Med 2018:doi:10.1002/mrm.27229.

19. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, Lian K, Kambhampati S, Kijowski R. Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage Lesion Detection. Radiology [Internet] 2018:172986. doi: 10.1148/radiol.2018172986.

20. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III Cham: Springer International Publishing; 2015 pp. 234–241. doi: 10.1007/978–3-319–24574-4_28.

21. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. ArXiv e-prints [Internet] 2015.

22. Guerra E, de Lara J, Malizia A, Díaz P. Supporting user-oriented analysis for multi-view domain-specific visual languages. Inf. Softw. Technol. [Internet] 2009;51:769–784. doi: 10.1016/j.infsof.2008.09.005.

23. Milletari F, Navab N, Ahmadi S-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. ArXiv e-prints [Internet] 2016.

24. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. ArXiv e-prints [Internet] 2016.

25. Wachinger C, Reuter M, Klein T. DeepNAT: Deep Convolutional Neural Network for Segmenting Neuroanatomy. ArXiv e-prints [Internet] 2017. doi: 10.1016/j.neuroimage.2017.02.035.

26. Dolz J, Desrosiers C, Ben Ayed I. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. Neuroimage [Internet] 2017. doi: 10.1016/J.NEUROIMAGE.2017.04.039.

27. Luc P, Couprie C, Chintala S, Verbeek J, Kuntzmann LJ. Semantic Segmentation using Adversarial Networks. Arxiv [Internet] 2016.

28. Kamnitsas K, Baumgartner C, Ledig C, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. ArXiv e-prints [Internet] 2016.

29. Moeskops P, Veta M, Lafarge MW, Eppenhof KAJ, Pluim JPW. Adversarial Training and Dilated Convolutions for Brain MRI Segmentation In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham; 2017 pp. 56–64. doi: 10.1007/978-3-319-6755897.

30. Kohl S, Bonekamp D, Schlemmer H-P, Yaqubi K, Hohenfellner M, Hadaschik B, Radtke J-P, Maier-Hein K. Adversarial Networks for the Detection of Aggressive Prostate Cancer. arXiv Prepr. arXiv1702.08014 [Internet] 2017.

31. Dai W, Doyle J, Liang X, Zhang H, Dong N, Li Y, Xing EP. SCAN: Structure Correcting Adversarial Network for Organ Segmentation in Chest X-rays. ArXiv e-prints [Internet] 2017.

32. Lafarge MW, Pluim JPW, Eppenhof KAJ, Moeskops P, Veta M. Domain-Adversarial Neural Networks to Address the Appearance Variability of Histopathology Images. In: Springer, Cham; 2017 pp. 83–91. doi: 10.1007/978-3-319-67558-9_10.

33. Greenspan Bram van HG, Summers RM, Greenspan H, et al. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. IEEE Trans. Med. Imaging [Internet] 2016;35:1153–1159. doi: 10.1109/TMI.2016.2553401.

34. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: 2017 IEEE International Conference on Computer Vision (ICCV) Vol. 2017– Octob. IEEE; 2017 pp. 2242–2251. doi: 10.1109/ICCV.2017.244.

35. Brislin RW. Back-Translation for Cross-Cultural Research. J. Cross. Cult. Psychol. [Internet] 1970;1:185–216. doi: 10.1177/135910457000100301.

36. Xia Y, He D, Qin T, Wang L, Yu N, Liu T-Y, Ma W-Y. Dual Learning for Machine Translation. ArXiv e-prints [Internet] 2016.

37. Zhou T, Krähenbühl P, Aubry M, Huang Q, Efros AA. Learning Dense Correspondence via 3D-guided Cycle Consistency. ArXiv e-prints [Internet] 2016.

38. Wang F, Huang Q, Guibas LJ. Image Co-segmentation via Consistent Functional Maps. In: 2013 IEEE International Conference on Computer Vision IEEE; 2013 pp. 849–856. doi: 10.1109/ICCV.2013.110.

39. Huang Q-X, Guibas L. Consistent Shape Maps via Semidefinite Programming. Comput. Graph. Forum [Internet] 2013;32:177–186. doi: 10.1111/cgf.12184.

40. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA. Context Encoders: Feature Learning by Inpainting. ArXiv e-prints [Internet] 2016.

41. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative Adversarial Text to Image Synthesis. ArXiv e-prints [Internet] 2016.

42. Denton E, Chintala S, Szlam A, Fergus R. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. ArXiv e-prints [Internet] 2015.

43. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. ArXiv e-prints [Internet] 2015.

44. Goodfellow IJI, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Networks. arXiv Prepr. arXiv … [Internet] 2014:1–9. doi: 10.1001/jamainternmed.2016.8245.

45. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. ArXiv e-prints [Internet] 2014;1411.

46. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. ArXiv e-prints [Internet] 2016.

47. Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J. Magn. Reson. Imaging [Internet] 2018. doi: 10.1002/jmri.25970.

48. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: Convolutional Architecture for Fast Feature Embedding. ArXiv e-prints [Internet] 2014;1408.

49. Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. ArXiv e-prints [Internet] 2016.

50. Li C, Wand M. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. ArXiv e-prints [Internet] 2016.

51. Heimann T, Morrison B, Styner M, Niethammer M, Warfield S. Segmentation of Knee Images: A Grand Challenge. Med Image Comput Comput Assist Interv 2010:207–214. doi: citeulike-articleid:12455897.

52. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. ArXiv e-prints [Internet] 2015;1502.

53. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. ArXiv e-prints [Internet] 2014.

54. Shan L, Zach C, Charles C, Niethammer M. Automatic atlas-based three-label cartilage segmentation from MR knee images. Med. Image Anal. [Internet] 2014;18:1233–1246. doi: 10.1016/j.media.2014.05.008.

55. Bron EE, van Tiel J, Smit H, Poot DHJ, Niessen WJ, Krestin GP, Weinans H, Oei EHG, Kotek G, Klein S. Image registration improves human knee cartilage T1 mapping with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC). Eur. Radiol. [Internet] 2013;23:246–52. doi: 10.1007/s00330-012-2590-3.

56. Chollet François. Keras. GitHub 2015:https://github.com/fchollet/keras.

57. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv e-prints [Internet] 2016. doi: 10.1109/TIP.2003.819861.

58. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? ArXiv e-prints [Internet] 2015.

59. Lekadir K, Galimzianova A, Betriu A, del Mar Vila M, Igual L, Rubin DL, Fernandez E, Radeva P, Napel S. A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound. IEEE J. Biomed. Heal. Informatics [Internet] 2017;21:48–55. doi: 10.1109/JBHI.2016.2631401.

60. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Adv. Neural Inf. Process. Syst. 25 [Internet] 2012:1097–1105.

61. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. ArXiv e-prints [Internet] 2016;1602.

62. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Trans. Med. Imaging [Internet] 2016;35:1299–1312. doi: 10.1109/TMI.2016.2535302.

63. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep Learning MR Imaging–based Attenuation Correction for PET/MR Imaging. Radiology [Internet] 2017:170700. doi: 10.1148/radiol.2017170700.

64. Jang H, Liu F, Zhao G, Bradshaw T, McMillan AB. Technical Note: Deep learning based MRAC using rapid ultra-short echo time imaging. Med. Phys. [Internet] 2018:In-press. doi: 10.1002/mp.12964.

65. Mirza M, Osindero S. Conditional Generative Adversarial Nets. ArXiv e-prints [Internet] 2014.

66. Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. ArXiv e-prints [Internet] 2017.

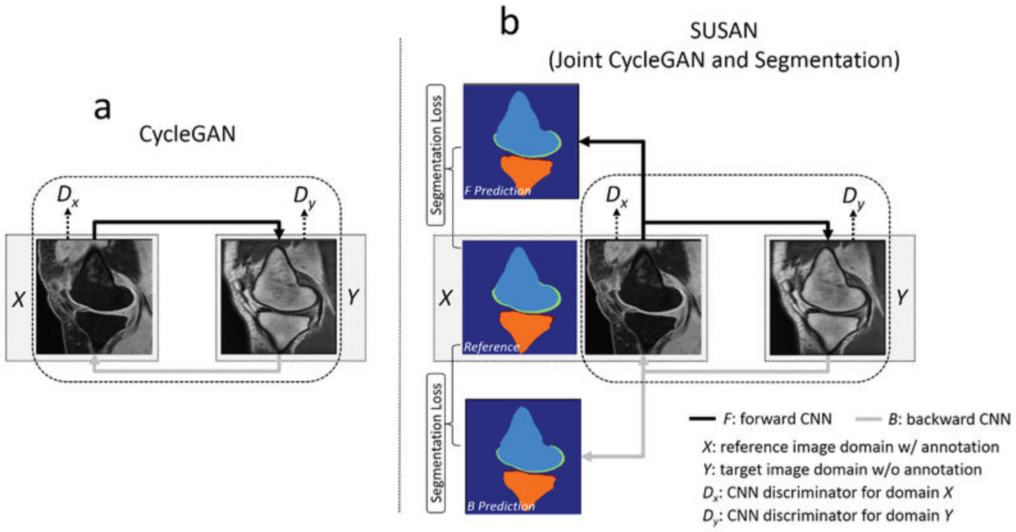
a
CycleGAN
D_x
D_y
X
Y
b
SUSAN
(Joint CycleGAN and Segmentation)
Segmentation Loss
F Prediction
Reference
B Prediction
Segmentation Loss
D_x
D_y
X
Y
F: forward CNN
B: backward CNN
X: reference image domain w/ annotation
Y: target image domain w/o annotation
D_x: CNN discriminator for domain X
D_y: CNN discriminator for domain Y

Figure 1.

a) The CycleGAN model contains two CNNs, F and B, for forward and backward mapping between image domain X (with annotation) and Y (without annotation), and associated adversarial discriminators D_X and D_Y. D_Y encourage F to translate images from domain X into outputs indistinguishable from domain Y, and vice versa for D_X and B. The cycle consistency loss is used to enforce the idea that if an image is translated from one domain to the other and back again, the image should look identical to its original version. **b)** SUSAN: the proposed joint CycleGAN and segmentation model incorporates an additional segmentation branch from the mapping CNNs. Joint training is performed for image translation between X and Y, and image segmentation for all image x from domain X and synthetic $F(x)$ which have origins from domain X using segmentation loss.

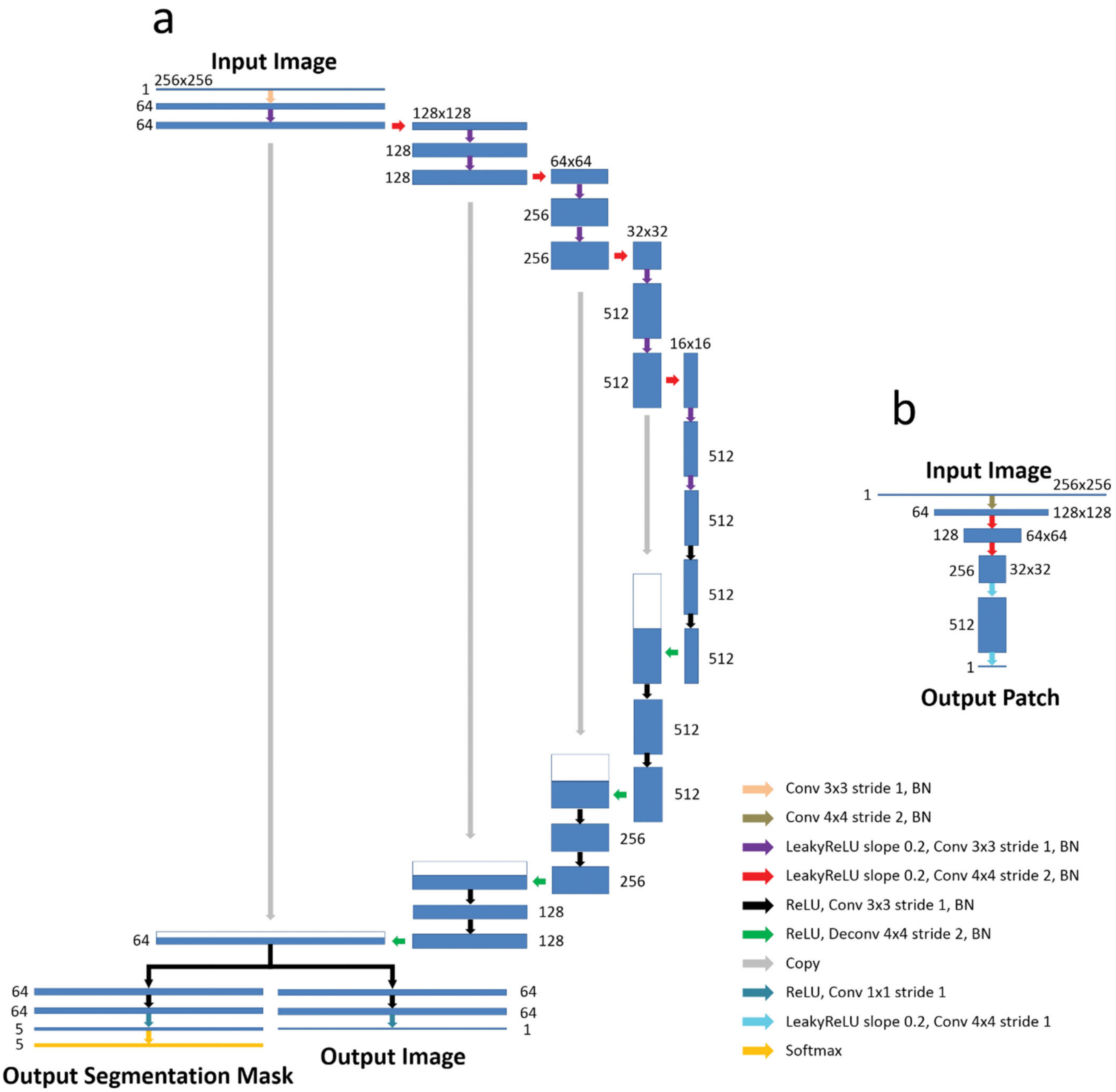

Figure 2.

A schematic illustration of the CNN architectures used in current SUSAN. **a)** The R-Net modified from standard U-Net structure is used for the CNN mapping function F and B. The R-Net allows two outputs each of which performs image segmentation and image translation, separately. The joint portion of the R-Net for segmentation and translation branch enables sharing image features during network training. **b)** The discriminator CNN designed in PatchGAN is used for D_X and D_Y. This network outputs image patches with reduced image size which will be used for differentiating real versus synthetic images in the adversarial training process. (Abbreviation: Conv: Convolution Layer; BN: Batch

Normalization; ReLU: Rectified Linear Unit; LeakyReLU: Leaky Rectified Linear Unit; Deconv: Transpose Convolution Layer; Softmax: Softmax Layer)

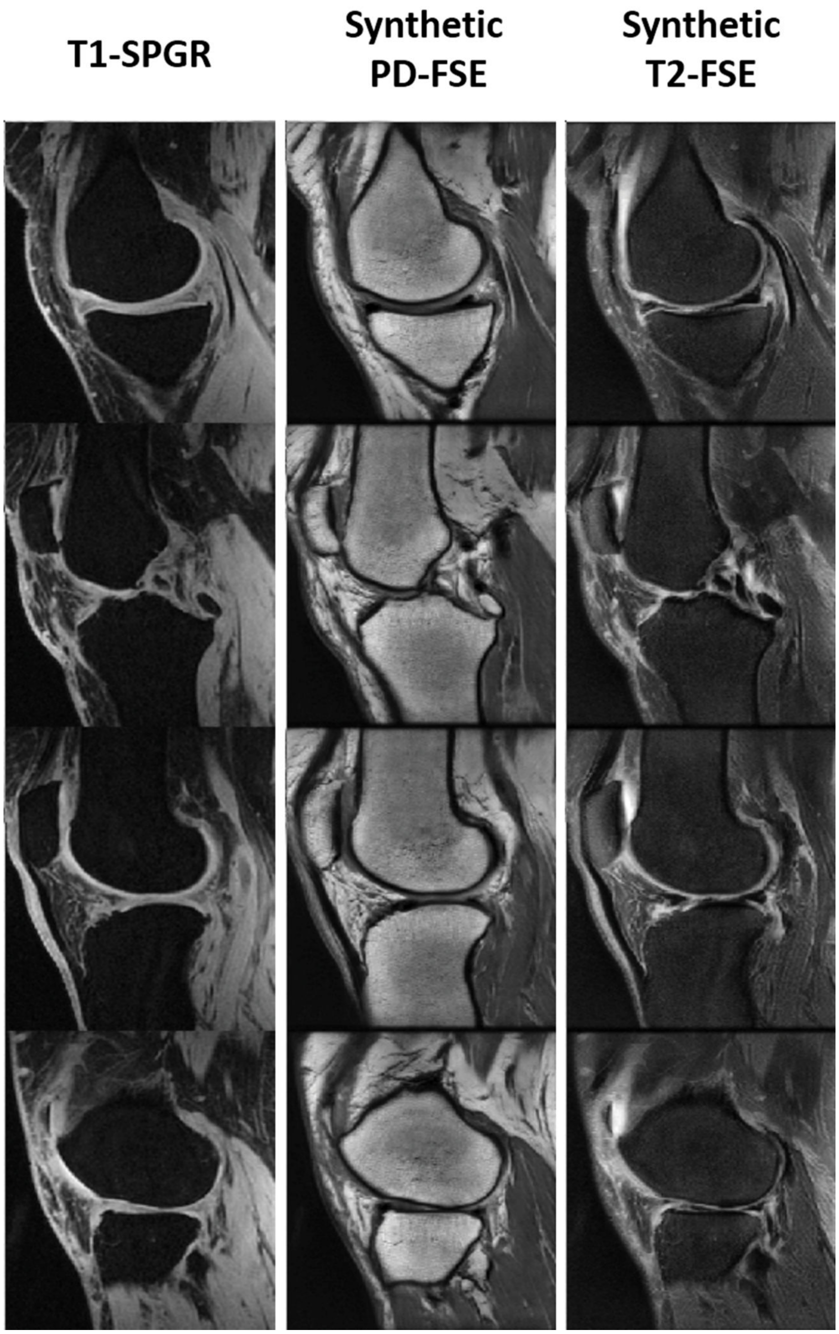

Figure 3.

Example of unpaired image-to-image translation for converting SKI10 T1-SPGR image contrast to the PD-FSE and T2-FSE image contrasts (denoted as synthetic PD-FSE and T2-FSE). Note that the subject was randomly selected for demonstration. Despite the dramatic differences in tissue contrasts between the three image datasets, SUSAN was capable of successfully translating the varying MR contrasts.

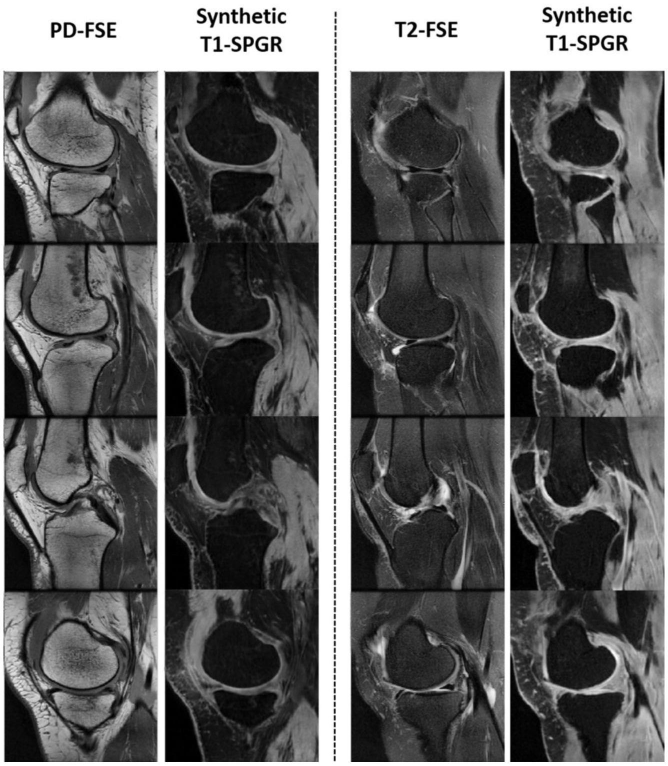

Figure 4.

Example of unpaired image-to-image translation for converting the PD-FSE and T2-FSE image contrasts to SKI10 T1-SPGR image contrast (denoted as synthetic T1-SPGR). Note that the subject was randomly selected from the individual image datasets for demonstration. Despite the dramatic differences in tissue contrasts between the three image datasets, SUSAN was capable of successfully translating the varying MR contrasts.

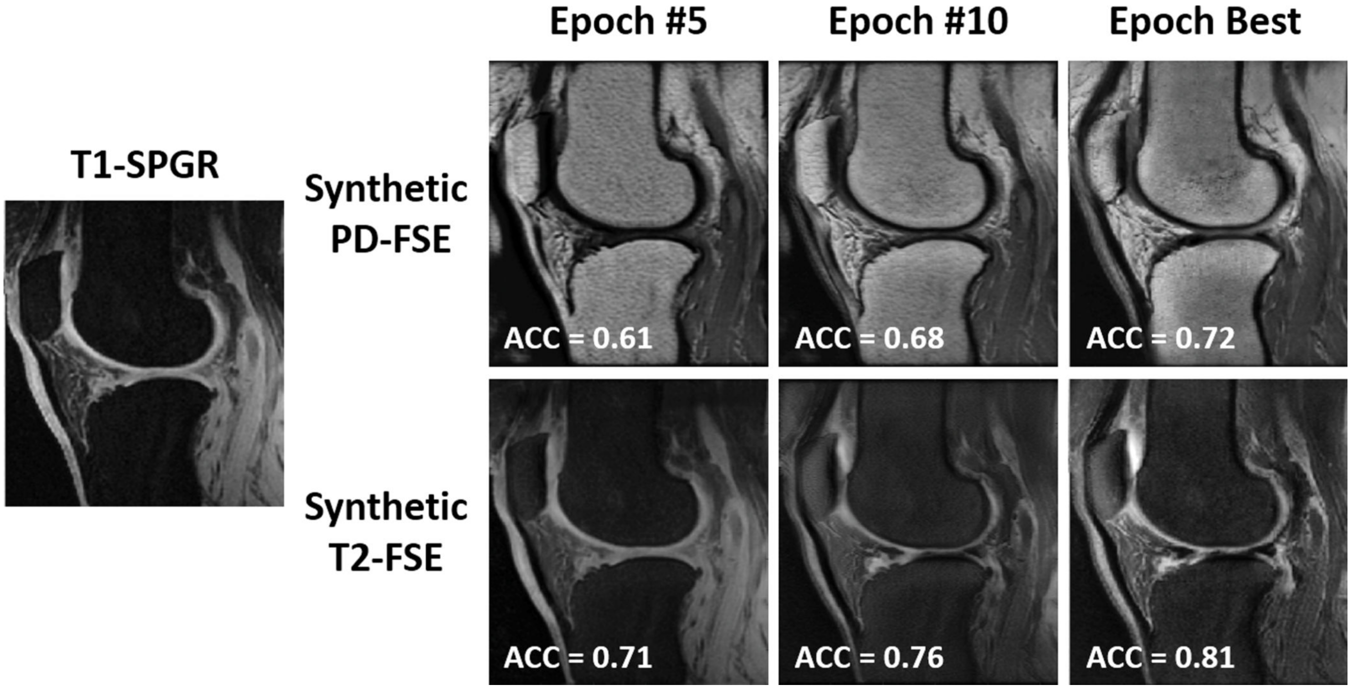

Figure 5.

Example of synthetic PD-FSE and T2-FSE images from a SKI10 T1-SPGR image at different epochs. The FCN-scores were calculated using ACC metric to illustrate quantitative assessment of the image quality.

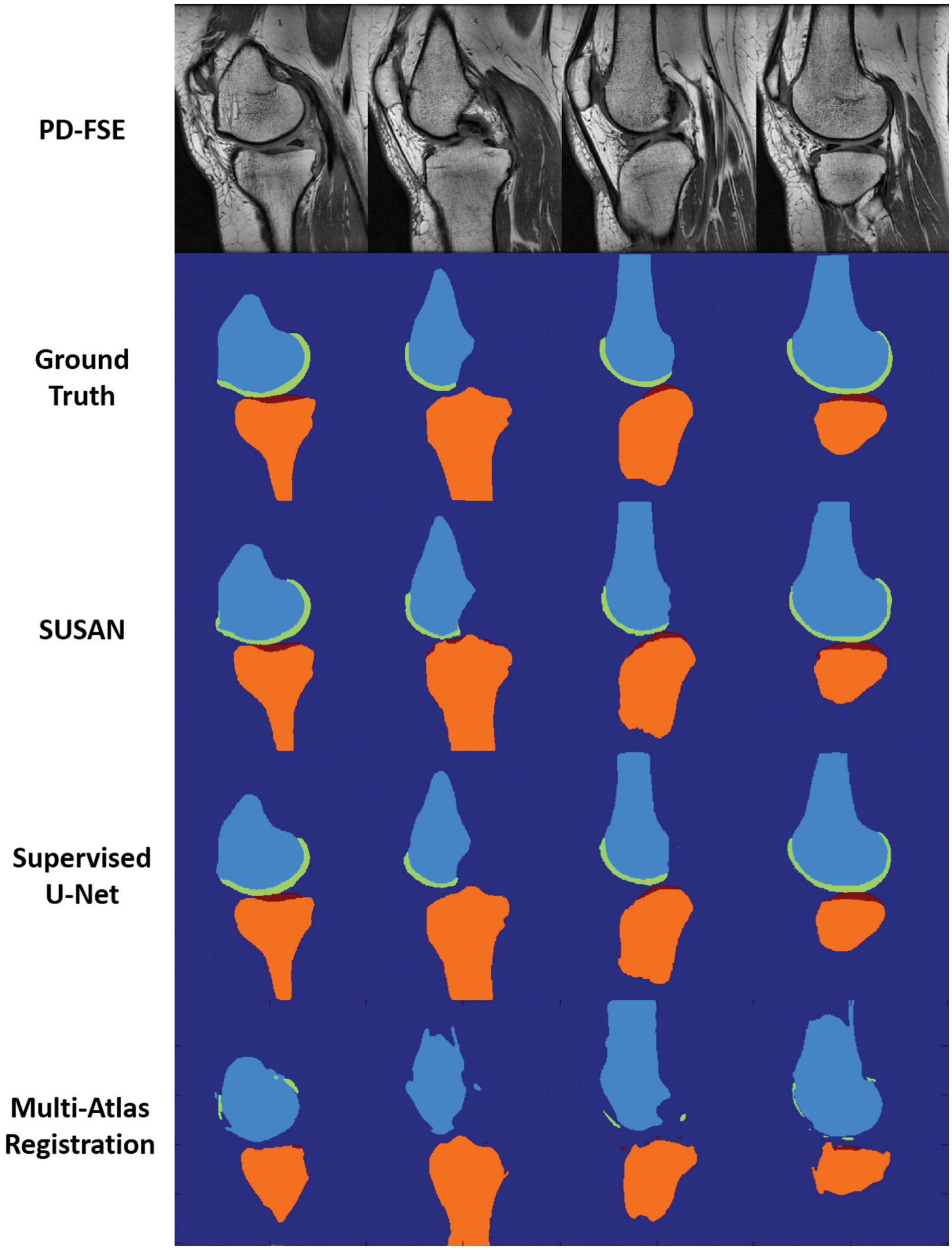

Figure 6.

Example of bone and cartilage segmentation for a 56 year old male subject with mild knee osteoarthritis performed on the PD-FSE image dataset. SUSAN provided accurate segmentation relative to the ground truth with segmentation performance comparable to the supervised U-Net method. Both deep learning methods outperformed the multi-atlas registration method.

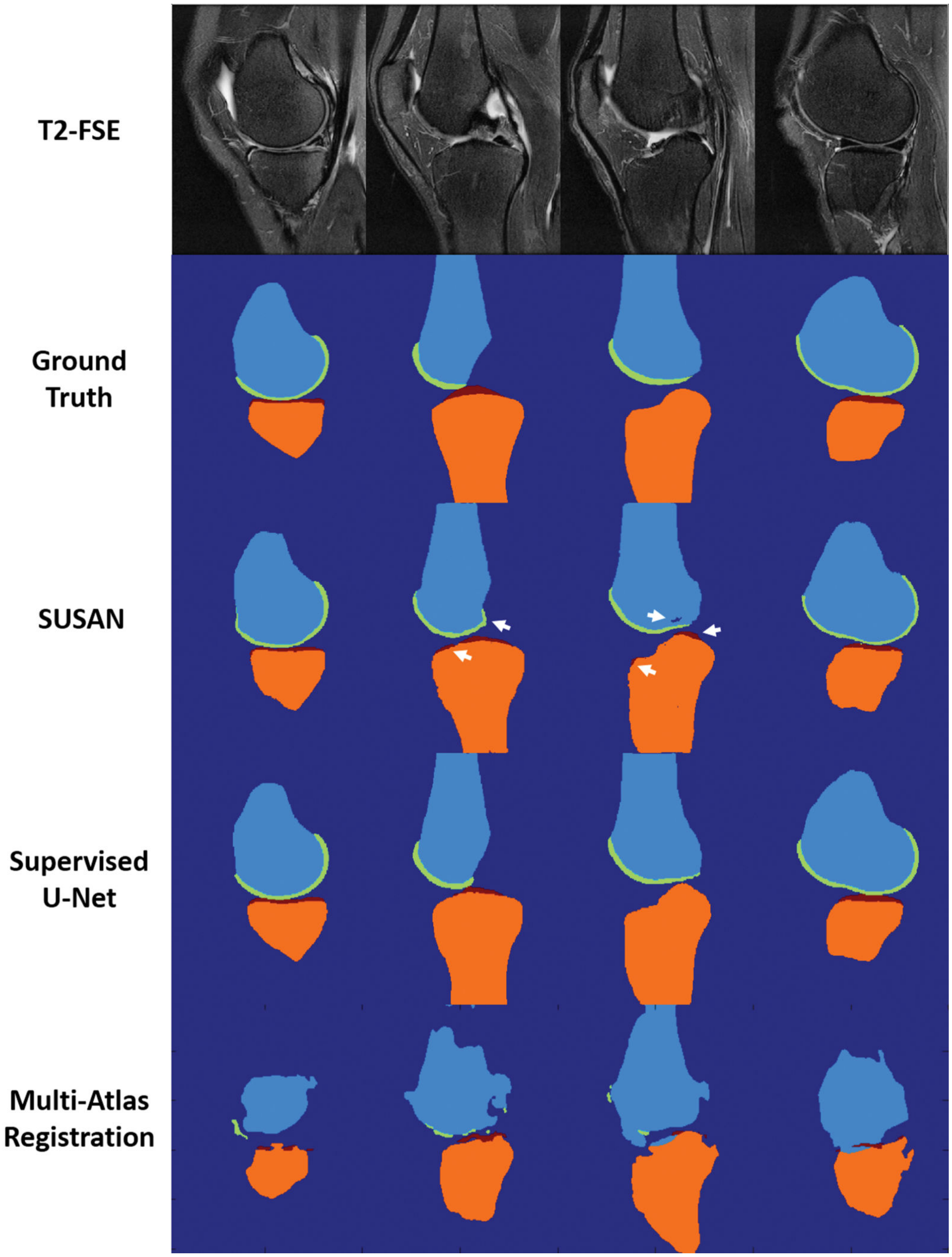

Figure 7.

Example of bone and cartilage segmentation for a 64 year old male subject with severe knee osteoarthritis performed on the T2-FSE image dataset. Despite a few clusters of misclassification for bone and cartilage (white arrows), SUSAN provided good overall contours relative to the ground truth with segmentation performance comparable to the supervised U-Net method. Both deep learning methods outperformed the multi-atlas registration method.

Table 1:

Imaging parameters for the sagittal fat-suppressed T1-weighted spoiled gradient-echo (T1-SPGR) sequence in the SKI10 image dataset and the two clinical MR sequences including sagittal fat-suppressed T2-weighted fast spin-echo (T2-FSE) and sagittal proton density-weighted fast spin-echo (PD-FSE).

Sequence	**T2-FSE**	**PD-FSE**	**T1-SPGR**
Field Strength (T)	3.0	3.0	1, 1.5, 3.0
TR [ms]	4680	1900	Unavailable
TE [ms]	80	19	Unavailable
Flip Angle [degree]	90	90	Unavailable
Pixel Bandwidth [Hz]	163	122	Unavailable
Echo Train Length	20	4	N/A
Field of View [cm]	14	14	14†
Slice Thickness [mm]	3	2	1
Slice Gap [mm]	0.5	0.5	N/A
Number of Slices	23~30	32~42	90~120
Matrix Size	512×512	512×512	360×300†

†Approximate values as actual matrix size and field of view vary with the size of the knee joint in the SKI10 image dataset

Table 2.

Comparison of image translation and segmentation accuracy (average value ± standard deviation) of DC for SUSAN at different weights of the segmentation branch for the two clinical MR image datasets.

Dataset	Weight of Segmentation Branch	FCN-Score	Dice Coefficient			
			Femur Bone	Tibia Bone	Femoral Cartilage	Tibial Cartilage
PD-FSE	0.5	0.71 ± 0.02	0.93 ± 0.01	0.93 ± 0.01	0.63 ± 0.02	0.62 ± 0.05
	5	**0.74 ± 0.03**	**0.97 ± 0.01**	**0.95 ± 0.00**	**0.66 ± 0.03**	**0.65 ± 0.06**
	10	0.73 ± 0.03	0.97 ± 0.01	0.93 ± 0.01	0.65 ± 0.03	0.64 ± 0.05
T2-FSE	0.5	0.75 ± 0.03	0.92 ± 0.01	0.90 ± 0.02	0.79 ± 0.02	0.71 ± 0.05
	5	**0.80 ± 0.05**	**0.95 ± 0.01**	**0.93 ± 0.02**	**0.81 ± 0.02**	**0.75 ± 0.06**
	10	0.78 ± 0.05	0.94 ± 0.00	0.93 ± 0.03	0.82 ± 0.02	0.75 ± 0.04

Table 3.

Comparison of segmentation accuracy (average value ± standard deviation) of DC, VOE and ASSD for SUSAN, the supervised U-Net method, and two registration methods for the two clinical MR image datasets. The asterisks were used to indicate the significant differences (*p*<0.05) of segmentation accuracy between the supervised U-Net method and SUSAN. The multi-atlas registration method performed significantly worse (*p*<0.001) than both SUSAN and the supervised U-Net method for all measures. The direct registration method also performed significantly worse (*p*<0.0001) than both SUSAN and the supervised U-Net method for all measures.

		Femur Bone		**Tibia Bone**		**Femoral Cartilage**			**Tibial Cartilage**		
Dataset	**Method**	**DC**	**ASSD [mm]**	**DC**	**ASSD [mm]**	**DC**	**VOE**	**ASSD [mm]**	**DC**	**VOE**	**ASSD [mm]**
PD-FSE	Supervised U-Net	0.96 ± 0.01	1.29 ± 0.31	0.96 ± 0.01	0.83 ± 0.43	*0.77 ± 0.01	*0.32 ± 0.03	0.50 ± 0.07	0.70 ± 0.04	0.36 ± 0.06	0.92 ± 0.21
	SUSAN	**0.97 ± 0.01**	**1.32 ± 0.25**	**0.95 ± 0.00**	**0.90 ± 0.16**	***0.66 ± 0.03**	***0.38± 0.05**	**0.57 ± 0.08**	**0.65 ± 0.06**	**0.38 ± 0.05**	**1.23 ± 0.32**
	Multi-Atlas Registration	0.86 ± 0.10	5.11 ± 2.11	0.80 ± 0.15	4.63 ± 3.15	0.25 ± 0.13	0.85 ± 0.11	5.16 ± 3.07	0.36 ± 0.21	0.82 ± 0.15	6.32 ± 2.46
	Direct Registration	0.71 ± 0.31	9.53 ± 3.0	0.65 ± 0.20	10.73 ± 5.32	0.15 ± 0.24	0.90 ± 0.10	10.2 ± 3.21	0.28 ± 0.15	0.85 ± 0.15	8.46 ± 6.22
T2-FSE	Supervised U-Net	0.96 ± 0.01	0.55 ± 0.27	*0.96 ± 0.02	*0.48 ± 0.24	0.82 ± 0.02	0.34 ± 0.03	0.59 ± 0.31	0.76 ± 0.03	0.31 ± 0.06	0.69 ± 0.43
	SUSAN	**0.95 ± 0.01**	**0.65 ± 0.20**	***0.93 ± 0.02**	***1.1 ± 0.28**	**0.81 ± 0.02**	**0.35 ± 0.04**	**0.65 ± 0.31**	**0.75 ± 0.06**	**0.33 ± 0.06**	**0.73 ± 0.36**
	Multi-Atlas Registration	0.75 ± 0.21	6.18 ± 3.31	0.70 ± 0.17	5.39 ± 4.09	0.31 ± 0.24	0.87 ± 0.17	7.18 ± 5.22	0.27 ± 0.16	0.89 ± 0.09	9.24 ± 4.93
	Direct Registration	0.65 ± 0.18	8.92 ± 4.46	0.53 ± 0.21	8.15 ± 4.39	0.26 ± 0.15	0.88 ± 0.18	8.29 ± 3.13	0.25 ± 0.14	0.89 ± 0.13	10.16 ± 4.52