https://doi.org/10.1142/9789819807154_0013

Chapter 1.13

LinAlign: X-Ray Image Alignment before and after Hip Arthroplasty

Cheng-Wei Lin*, Alexander Yurusov, Hsiao-Ning Chen, and Chiou-Shann Fuh

Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
**r10922102@ntu.edu.tw*

In this chapter, we propose LinAlign[8]: a computer vision algorithm for aligning X-ray images before and after surgery, providing a system for surgeons to compare images before and after surgery more efficiently and replace manually aligning procedure. LinAlign allows only align specific areas when aligning images and solves the problem that linear transformation cannot be performed on non-rigid objects. Therefore, it is suitable for comparing the position of bones during hip replacement surgery, allowing orthopedic surgeons to make sure that implants have been installed correctly.

In our experiment, we took the X-ray images of the pelvis as our experimental data: Each set of images contains the X-ray photographs of the same patient taken at different times. We experiment with different methods. By comparing similar features between images and calculating the displacement of these feature points, the images can be aligned.

We evaluate the performance of the algorithm by the error of pre-defined landmarks after alignment. These landmarks are anatomically important features of the skeletal system. The goal of our experiment is to minimize the distance of landmarks between image pairs. We take the mean square error of these landmark distances as the performance metric of our algorithm.

1. Introduction

In orthopedic surgery, surgeons will take an X-ray image and compare it with the image taken before surgery to ensure the surgery is accurate. We aim to develop a system to align these images. When surgeons need to compare the current image with the image taken before surgery, our system can align the previous one with the current one automatically. Therefore, they no longer need to print the hardcopy of radiographs and compare them manually.

Our research focuses on the radiograph taken from Total Hip Arthroplasty (THA), which is a surgical procedure where the hip joint is replaced by a prosthetic

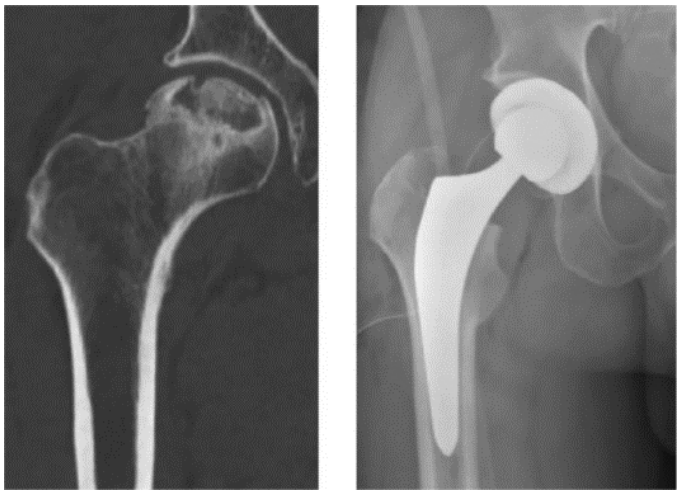


Fig. 1. An X-ray image showing a left hip has been replaced with prosthetic implant due to advanced femoral head collapse.[12]

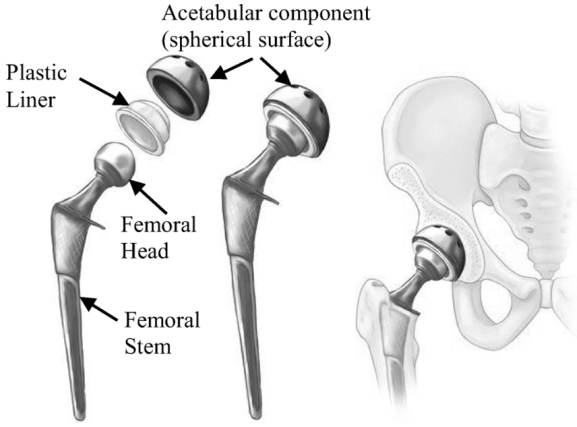


Fig. 2. The components of prosthetic implant for total hip replacement.[11]

implant. Figure 1 shows the radiographs taken before and after THA, and Figure 2 shows the components of the implant used in THA.

Total hip arthroplasty is a major surgery that carries certain risks. This includes infection, dislocation, limb length inequality, fracture, etc. Dislocation is the most common complication in this surgery (see Figure 3). For example, the ball coming out of the socket. It may be caused by misalignment of the prosthetic implant and hip joint, which can be reduced by ensuring the location of them is correct during the surgery. Our system is designed to reduce these risks.

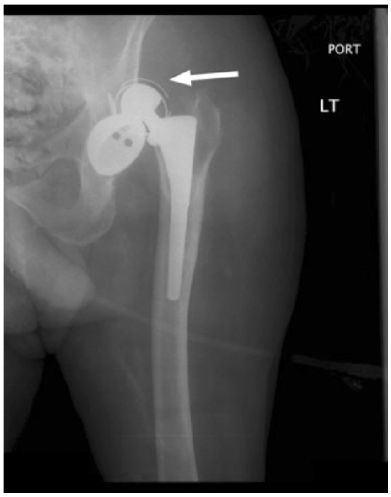


Fig. 3. The dislocation of the femoral head.[2]

2. Related Work

2.1. *Feature matching*

Feature matching in computer vision is the process of identifying and comparing the same or similar features in two or more images. Often, the matching result contains several wrong correspondences, which can easily affect our application. To prevent the effect of outlying data, a noise removal phase such as RANSAC (RANdom SAmple Consensus)[6] is usually applied at the end. With this process, the noisy matching can be almost removed.

Depending on the approach, feature matching methods can be categorized into two types: detector-based feature matching and detector-free feature matching. Detector-based methods require a feature detector to extract and describe local features, and then match these feature points. As for detector-free methods, they remove the feature detector phase and directly produce dense descriptors or dense feature matches instead.

Feature matching can be used in various computer vision tasks, such as image registration, object recognition, Simultaneous Localization And Mapping (SLAM), and Structure-from-Motion (SfM).

2.2. *Homography estimation*

Homography estimation is a computer vision task that finds a mathematical relationship (i.e., a homography matrix) between multiple images taken from different perspectives. The homography matrix estimation is based on corresponding points between two images and can be computed mathematically, such as direct linear transformation and normalized direct linear transform. Homography estimation can

be used in computer vision tasks, such as image alignment, image stitching, image mosaicing, and object recognition.

2.3. *Semantic segmentation*

Semantic segmentation is a computer vision task assigning the semantic label to each pixel in the image. The goal is to categorize each pixel in the image into one of several predefined classes. The result of semantic segmentation is a dense prediction, where every pixel in the image is assigned a color-coded class label. Semantic segmentation is a crucial task for various applications such as self-driving cars which can make the decisions informed by its fine-grained prediction of the environment.

Recently, deep neural networks have achieved successful performance in semantic segmentation, such as Fully Convolutional Networks (FCN),[8] SegNet,[1] U-Net,[13] DeepLab,[3] and so on. The networks take an image as input and output a corresponding segmentation mask where each pixel is assigned a class label.

2.4. *Image alignment*

Image alignment is the process of matching and adjusting the relative positions of images to align them with each other. By the matching of common features between multiple images, we can align the images so that the same feature in different images corresponds to the same real-world location.

There are two common methods for image alignment: feature point-based alignment and optical flow alignment. Feature point-based alignment aims to align the feature point correspondences in multiple images by calculating the global transformation matrices between them. Since it is a linear transformation, it can only perform the translation, rotation, and scaling on a rigid body. In other words, the motion in images is restricted to be globally uniform. As for optical flow alignment, it aligns images by remapping with dense correspondences. Since it is a one-to-one correspondence between pixels, images can be aligned accordingly, but its result is often unstable.

3. Methodology

We propose LinAlign for aligning X-ray images before and after surgery. Since the pose in X-ray images may be different and cannot be aligned linearly, we propose a new strategy to compute homography to deal with this issue.

For an X-ray image pair, we first perform the feature matching to compute the point correspondences between both images. Simultaneously, we compute the probability map of the pelvis with a semantic segmentation model, which results in images with pixel-wise class prediction. Then, we assign the weight to each point correspondence with the product of class confidence on both images and perform

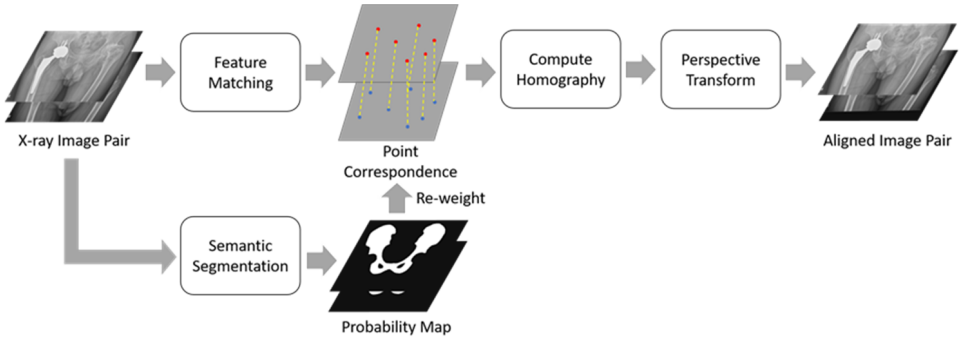


Fig. 4. The pipeline of LinAlign algorithm.[8]

the homography estimation using weighted normalized Direct Linear Transform (DLT) on these correspondences. Finally, we align both images with perspective transformation. Figure 4 shows the pipeline of LinAlign algorithm.

3.1. *Feature matching*

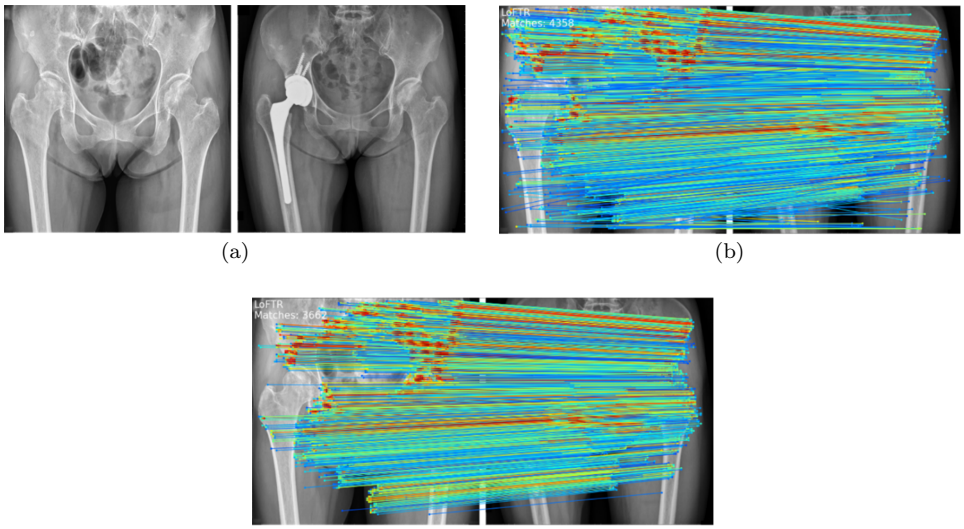
We use LoFTR[15] to match the feature points from two images. LoFTR predicts thousands of dense point correspondences, and it works even in low-texture areas. These correspondences are accurate and robust, so they give us good information for image alignment.

Our LoFTR model was pre-trained on MegaDepth[7] dataset, which contains a large amount of outdoor scene images collected from the Internet with depth map annotations. LoFTR learns how to match the dense features directly, so it can be easily adapted to other domains, even if the scene is very different from training datasets. Thus, it achieves good performance in our application.

After feature matching, some outlying correspondences may exist. Since our goal is to align the images, these outlying data can seriously affect our results, even with only a small number of outliers. To prevent this effect, we apply a noise removal phase using RANSAC at the end. Thus, we can select the good matchings that are helpful to align the images. Figure 5 shows the feature-matching prediction of LoFTR tested on our data. Applying the RANSAC can remove the correspondences that are outlying to the linear transform.

3.2. *Pelvis segmentation*

In our application, we aim to compare the location of the hip joint before and after surgery. Since the human body is non-rigid, it may have pose difference and cannot align the whole body with linear transformation. However, in this application, we only need to align the pelvis part well, which is the rigid part. Therefore, in this



(a) (b)

(c)

Fig. 5. The feature matching results with LoFTR.[15] (a) Original image pair. (b) Matching result. (c) Matching result with outlier removal using RANSAC.

procedure, we segment the pelvis part from images, and use it to improve the alignment.

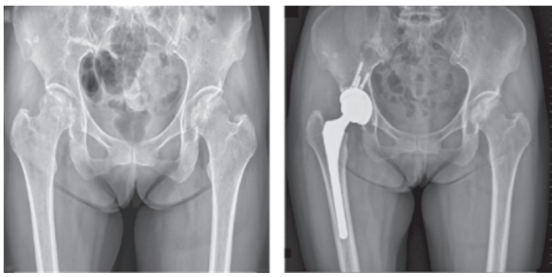
We segment the pelvis part with U-Net++,[17] which is an improved version of U-Net.[13] We use X-ray images of the pelvis to train our model, with 191 images as training data and 33 images as validation data. Pixels in each image are labeled with two classes, where label 0 is the background and label 1 is the pelvis.

The model outputs a probability map with the size (N, H, W), where N is the number of classes and H and W are the height and width of the input image. Each pixel represents the class prediction. We take the probability prediction of the pelvis as our result to support our image alignment. Figure 6 shows the probability map prediction on our dataset using our U-Net++ model. The pelvis area can be segmented perfectly and so is helpful to our algorithm.

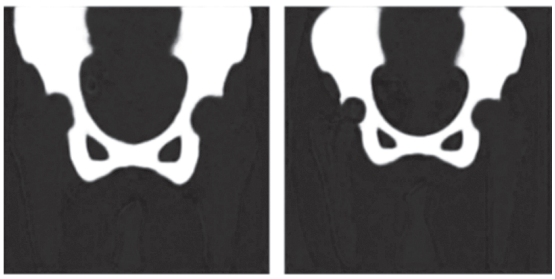
3.3. *Weighted normalized DLT*

We propose a new method to estimate homography from point correspondences: weighted normalized DLT (Direct Linear Transform). Since the goal of our algorithm is to compare two images taken before and after surgery, it would be more suitable for comparison if we align the pelvis area better. Therefore, we modify the previous homography estimation methods to fit our requirements.

Our method is done by assigning the weight to each correspondence while computing the normalized DLT. Therefore, the correspondence with higher weight can be more likely to be aligned.



(a)



(b)

Fig. 6. The probability map prediction result of our data.

3.3.1. *DLT*

Assume that we are given N correspondence key point pairs, where the point correspondences are $\{(u_i, v_i), (u'_i, v'_i) | i = 1, \ldots, N\}$, λ_i is a point-dependent scale factor, we have

$$\lambda_i \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} u'_i \\ v'_i \\ 1 \end{bmatrix}.$$

Then, we find the homography matrix H by replacing the equation with matrix form:

$$\begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -u'_1 u_1 & -u'_1 v_1 & -u'_1 \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -v'_1 u_1 & -v'_1 v_1 & -v'_1 \\ u_2 & v_2 & 1 & 0 & 0 & 0 & -u'_2 u_2 & -u'_2 v_2 & -u'_2 \\ 0 & 0 & 0 & u_2 & v_2 & 1 & -v'_2 u_2 & -v'_2 v_2 & -v'_2 \\ & & & & \vdots & & & & \\ & & & & \vdots & & & & \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} = 0$$

$$\rightarrow A\boldsymbol{h} = 0.$$

Since the Degree of Freedom (DoF) for perspective transformation is 8, we need at least 4 correspondences to find the non-trivial solution to this equation. We thus solve the least-squared-error solution:

$$\min_{\boldsymbol{h}} \|A\boldsymbol{h}\|^2, s.t.\ \|\boldsymbol{h}\|^2 = 1.$$

The solution is equivalent to

$$\underset{\boldsymbol{h}}{argmin} \frac{\|A\boldsymbol{h}\|^2}{\|\boldsymbol{h}\|^2} = \frac{h^T A^T A h}{\|\boldsymbol{h}\|^2}.$$

According to the principle of Rayleigh quotient, the solution of $\boldsymbol{h}$ is the eigenvector corresponding to the smallest eigenvalue of matrix $A^T A$. Thus, the solution can be obtained by performing the Singular Value Decomposition (SVD) on A, so we get $A = U\Sigma V^T$, where the singular values in Σ are arranged in descending order. The solution is then the last column of V.

3.3.2. *Normalized DLT*

There is a shortcoming in DLT. When we compute the linear transform, the x, y coordinate on the image is typically much larger than z coordinate since z is always equal to 1. The different order of coordinates results in an ill-conditioning number of matrix A and numerically unstable solutions. To get more stable results, we can normalize every point before direct linear transformation. Assume T, T' are the transformation matrices that normalize the points $(u_i, v_i), (u'_i, v'_i)$ respectively:

$$T = \begin{bmatrix} s & 0 & m_x \\ 0 & s & m_y \\ 0 & 0 & 1 \end{bmatrix}^{-1},$$

where s is the standard deviation of coordinate and m_x, m_y are the mean of x-axis and y-axis. After performing the direct linear transformation on normalized correspondence point, we get $\widehat{H}$. The transformation matrix between (u_i, v_i) and (u'_i, v'_i) will be $H = T'^{-1}\widehat{H}T$. Thus,

$$\lambda_i \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \approx \begin{bmatrix} u'_i \\ v'_i \\ 1 \end{bmatrix}.$$

3.3.3. *Weighted normalized DLT*

In previous methods, point correspondences have the same weight during the homography estimation. However, in our application, we want our interested area (i.e., pelvis area) more likely to be aligned, so we can compare the images better.

Weighted Normalized DLT is done by modifying the equation of Normalized DLT. We re-weight the rows in A, so we get

$$\begin{bmatrix} w_1 \begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1'u_1 & -u_1'v_1 & -u_1' \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -v_1'u_1 & -v_1'v_1 & -v_1' \end{bmatrix} \\ w_2 \begin{bmatrix} u_2 & v_2 & 1 & 0 & 0 & 0 & -u_2'u_2 & -u_2'v_2 & -u_2' \\ 0 & 0 & 0 & u_2 & v_2 & 1 & -v_2'u_2 & -v_2'v_2 & -v_2' \end{bmatrix} \\ \vdots \\ \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} = 0$$

$$\rightarrow A'\widehat{\boldsymbol{h}} = 0.$$

We solve $A'\widehat{\boldsymbol{h}} = 0$ in the same approach as we described in Section 3.3.1, we get $\widehat{H}$. Then, we denormalize it to get the homography matrix $H = T'^{-1}\widehat{H}T$. Thus, we have

$$\lambda_i H \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \approx \begin{bmatrix} u_i' \\ v_i' \\ 1 \end{bmatrix}.$$

3.4. *Image alignment*

We align the images using perspective transformation. For given point correspondences and the probability map of the pelvis area determined by the segmentation model, we tried two methods to compute the homography matrix.

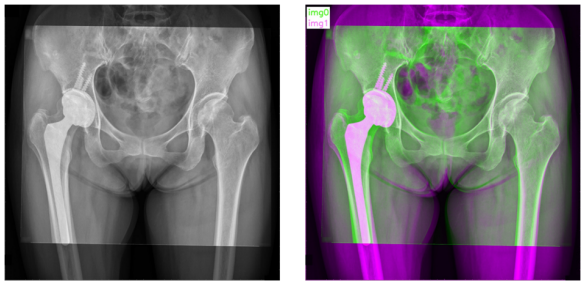
Let P_c^1, P_c^2 be the probability map prediction of class c from the semantic segmentation model and

$$M = \{(u_i, v_i), (u_i', v_i') | i = 1, \ldots, N\}$$

be our keypoint correspondences. In Method 1, we filter the keypoints with the label prediction and compute the normalized DLT with the keypoints within the pelvis area only. Take

$$\big\{(u_i, v_i), (u_i', v_i') | P_{pelvis}^1(u_i, v_i) \geq 0.5 \quad \text{and} \quad P_{pelvis}^2(u_i, v_i) \geq 0.5\big\}$$

to compute normalized DLT; as for Method 2, we assign the weight with the product of the confidence score of the pelvis from both images to compute weighted normalized DLT, i.e., $w_i = P_{\text{pelvis}}^1(u_i, v_i) \times P_{\text{pelvis}}^2(u_i', v_i')$. Testing these methods on our datasets and comparing them with the normalized DLT method, we finally take Method 2 because it has better performance.



(a) (b)

Fig. 7. The visualization of our aligning results. (a) The overlapping image. (b) The color difference map (CDM).

3.5. *Result visualization*

In order to make surgeons easier to compare the images using alignment result, we visualize the alignment using CDM (Color Difference Map), implemented based on YCbCr color space, where Y is the luminance, and Cb and Cr are the blue-difference and red-difference chroma components.

Assume that I_1, I_2 is our image pair, where I_2 is our target image. These images are grayscale images. We align I_1 with I_2, so we get a transformed image I_1'. The CDM is determined by assigning overlapping image to Y and color difference to Cr and Cb, i.e., $\mathrm{Y} = (I_1' + I_2)/2$; $\mathrm{Cr} = (I_2 - I_1')/2$; $\mathrm{Cb} = (I_2 - I_1')/2$ in Fig. 7.

4. Experimental Results

4.1. *Datasets*

We experiment with a dataset of pelvis radiographs which contains 224 radiographs collected from 49 patients.

In our application, we require semantic segmentation dataset for model training and pelvic landmarks dataset for performance evaluation. We create these by labeling our first dataset with a pelvis mask and landmark point set. To test the repeatability, we experiment with our algorithm on both datasets. Figure 8 shows a part of our segmentation dataset. Figure 9 shows the definition of landmarks.

4.2. *Feature matching experiment*

Feature matching is a crucial step in our application. To improve our alignment, we experiment with three feature-matching methods: SIFT,[10] SuperPoint[5] + SuperGlue,[14] and LoFTR.[15]

We evaluate the alignment performance using MRE (Mean Radial Error). Given keypoint correspondences $\{(u_i^1, v_i^1), (u_i^2, v_i^2) | i = 1, \ldots, N\}$ as our ground truth, where $(u_i^1, v_i^1), (u_i^2, v_i^2)$ are the pre-defined landmarks on both images. For every

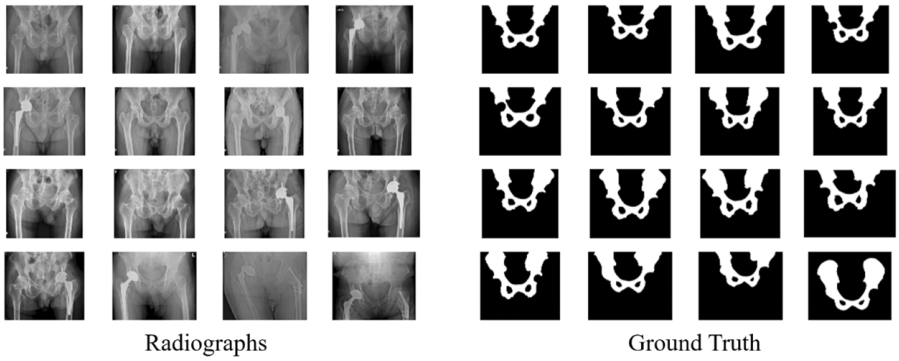
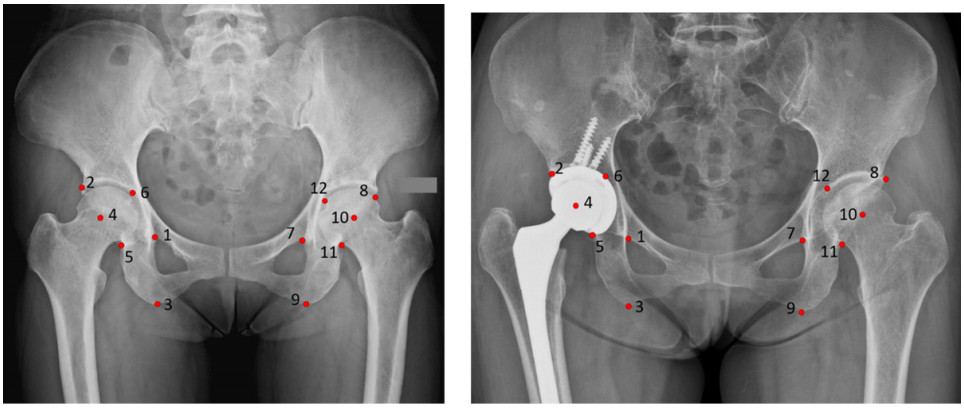


Fig. 8. Samples of our segmentation dataset containing pelvis radiographs and the mask annotations. Some of them are taken after THA surgery.



(a) (b)

Fig. 9. Definition of the landmarks. (a) Normal. (b) With implants.

keypoint pair, calculate the distance between $(u_i^1, v_i^1)'$ and (u_i^2, v_i^2), where $(u_i^1, v_i^1)'$ is the projection of (u_i^1, v_i^1). Thus, the error is defined to be

$$MRE = \frac{1}{N} \sum_{i=1}^{N} \left\| (u_i^1, v_i^1)' - (u_i^2, v_i^2) \right\|_2 .$$

We perform the homography test on our landmark dataset by estimating the homography using RANSAC.[6] LoFTR gets the best result finally, so we select LoFTR in our LinAlign algorithm. Table 1 shows the homography estimation results.

4.3. *Semantic segmentation experiment*

We experiment with four state-of-the-art semantic segmentation models: U-Net,[13] U-Net++,[17] DeepLabV3+,[4] and PSPNet,[16] tested on our segmentation dataset.

Table 1. Homography estimation results.

Methods	**SIFT**	**SP+SG**	**LoFTR**
MRE ↓	0.134639	0.015056	**0.014456**

Note: ↓: The lower, the better.

Table 2. Semantic segmentation test on our validation set.

Model	**U-Net**	**U-Net++**	**DeepLabv3+**	**PSPNet**
IoU Score ↑	0.9233	**0.9314**	0.9246	0.7236

Note: ↑: The higher, the better.

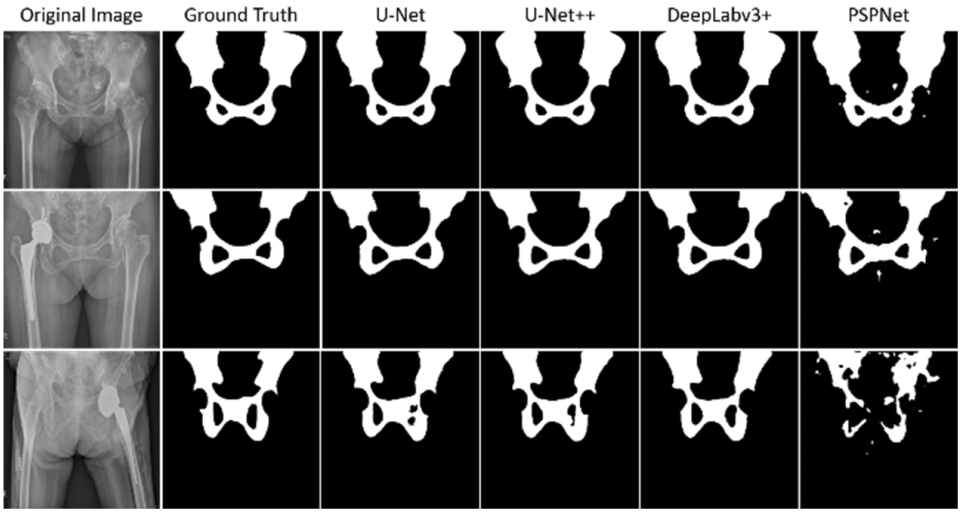


Fig. 10. Segmentation results on validation set.[8]

These models are trained in the same conditions. They trained with 64 epochs, supervised with pixel-wise cross-entropy loss, and the data augmentation strategies are random horizontal flip, random shift-scale rotation, random brightness contrast, and random resized crop. We take the IoU (Intersection over Union) score of pelvis area as our performance, where

$$IoU = \frac{TP}{TP + FP + FN}.$$

Table 2 shows the IoU score on our validation data. U-Net++ is the best model in our experiment. Figure 10 shows the segmentation results on validation set for these models with their ground truth.

Table 3. Image alignment testing results.

Method	Normalized DLT	Filtered	Re-weighted
MRE ↓	0.014470	0.014165	**0.013985**

Note: ↓ : The lower, the better.

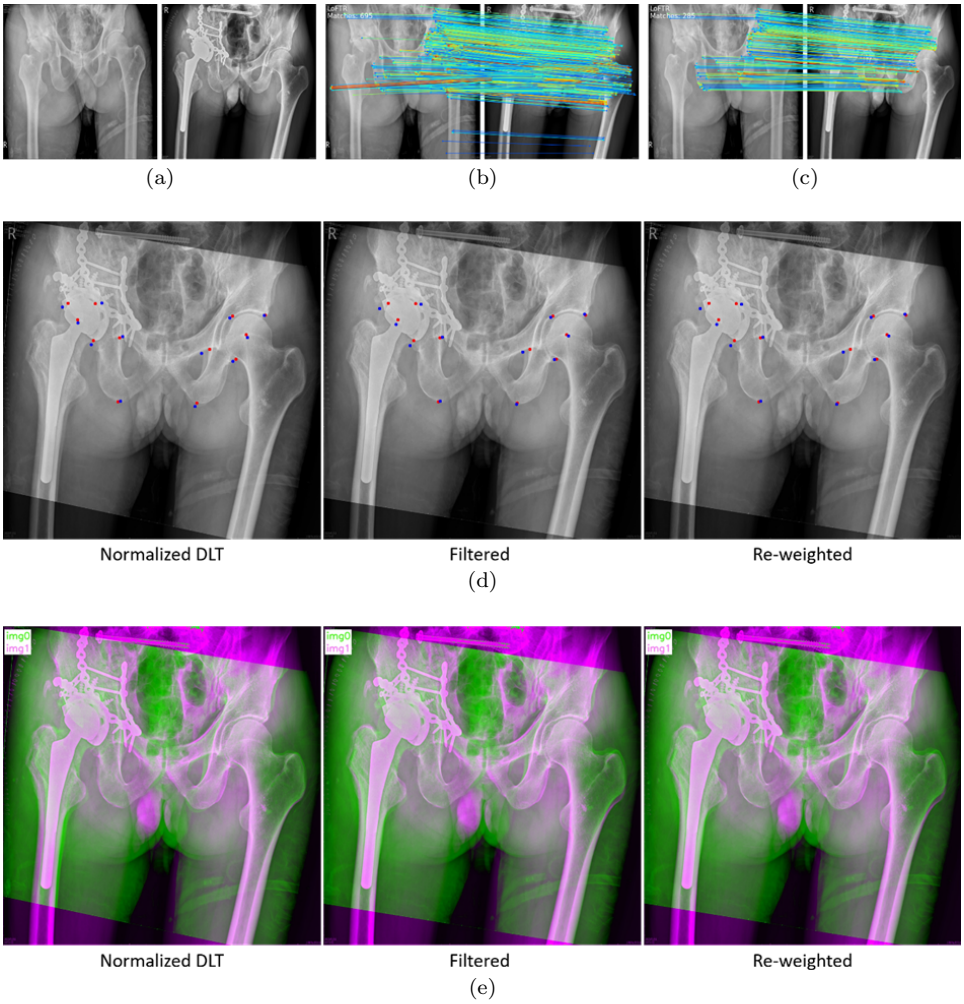


Fig. 11. Alignment result 1. (a) Original image pair. (b) Matching result. (c) Matches within pelvis area. (d) Alignment results with landmarks. (e) CDM. (The MRE are 0.013362, 0.012644, and 0.012767, respectively.)[8]

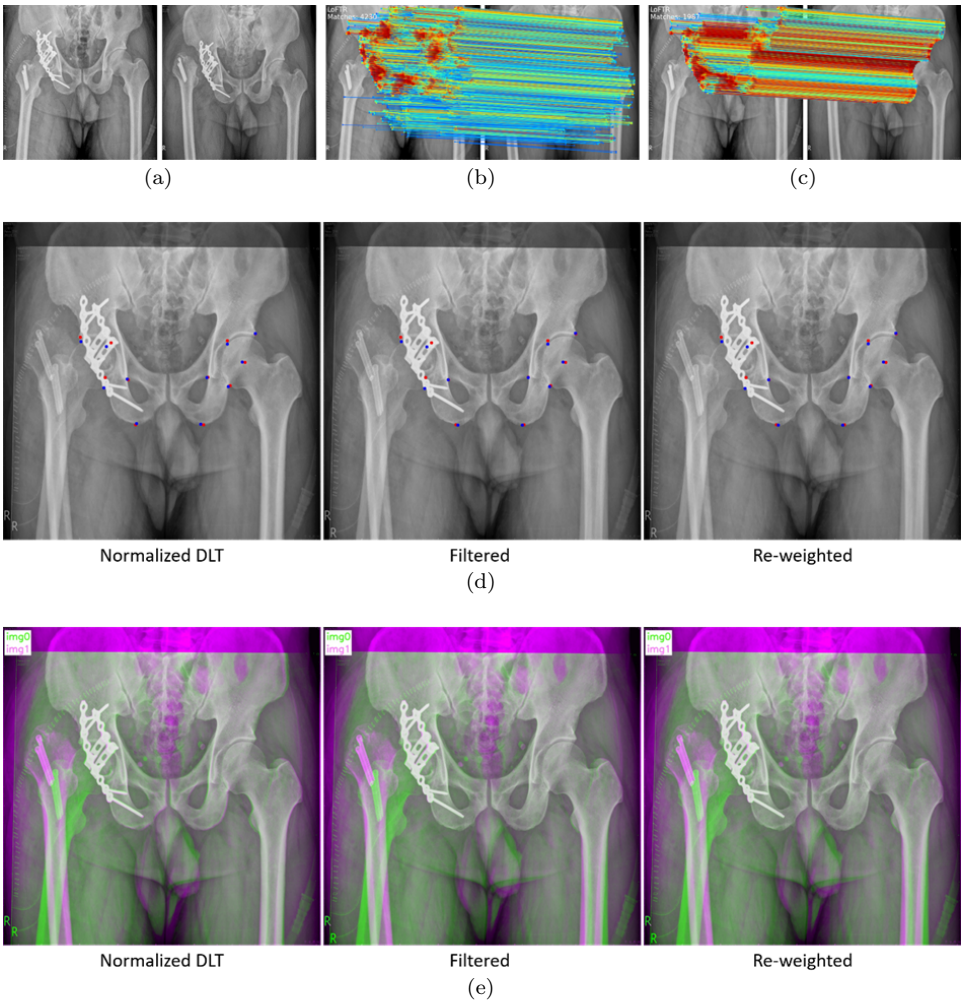


Fig. 12. Alignment result 2. (a) Original image pair. (b) Matching result. (c) Matches within pelvis area. (d) Alignment results with landmarks. (e) CDM. (The MRE are 0.01024, 0.009574, and 0.009485, respectively.)[8]

4.4. *Image alignment experiment*

We experiment with both proposed methods in Section 3.3 and normalized DLT on our landmarks dataset to compare the performance, using MRE of landmarks as our evaluation protocol. Tested with i7-9770 and RTX 2080ti, the inference time for an image pair is 2.23 seconds. Table 3 shows the MRE scores of these methods. Figures 11–15 show the experimental results. In (d), red points are landmarks

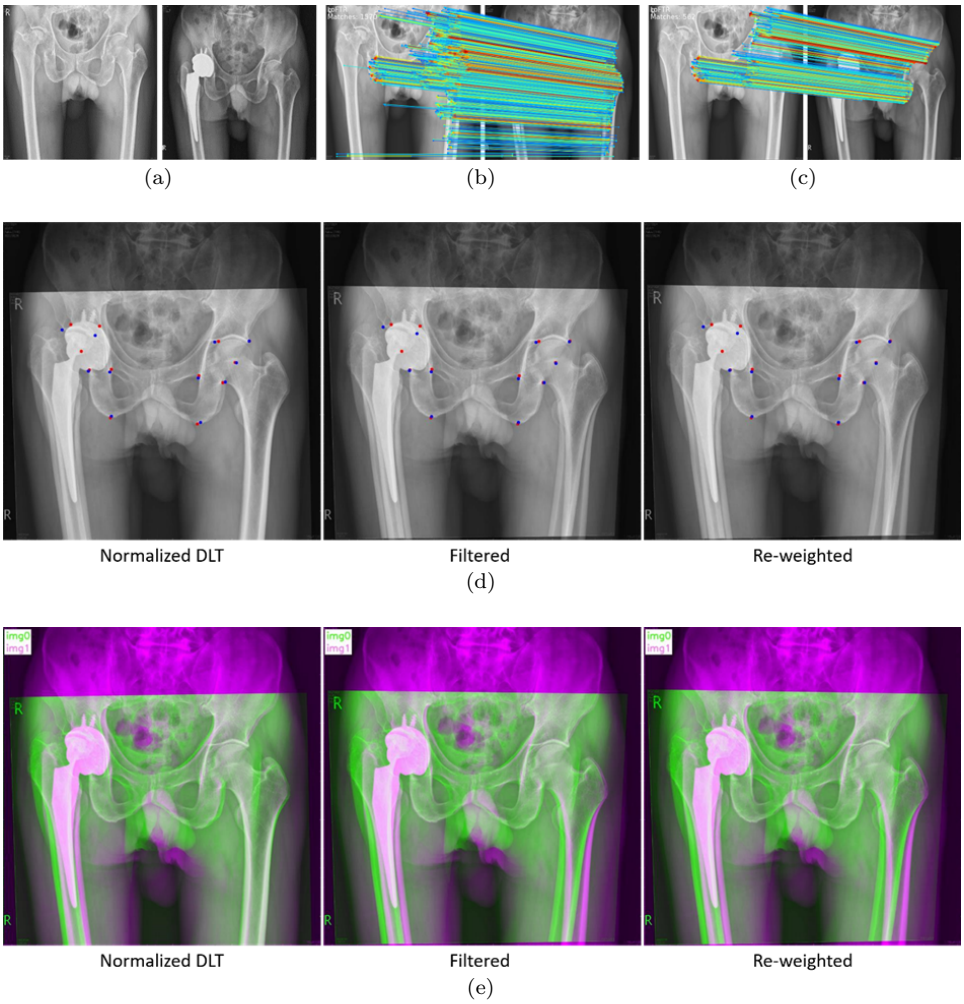


Fig. 13. Alignment result 3. (a) Original image pair. (b) Matching result. (c) Matches within pelvis area. (d) Alignment results with landmarks. (e) CDM. (The MRE are 0.011629, 0.009012, and 0.008951, respectively.)[8]

on the transformed image, and blue points are landmarks on the target image. Extensive experiments show that our algorithm can align most of the images well. However, due to the complexity of the 3D structure of the pelvis, it is difficult to handle the view difference, but we still can align the joint part as well as possible.

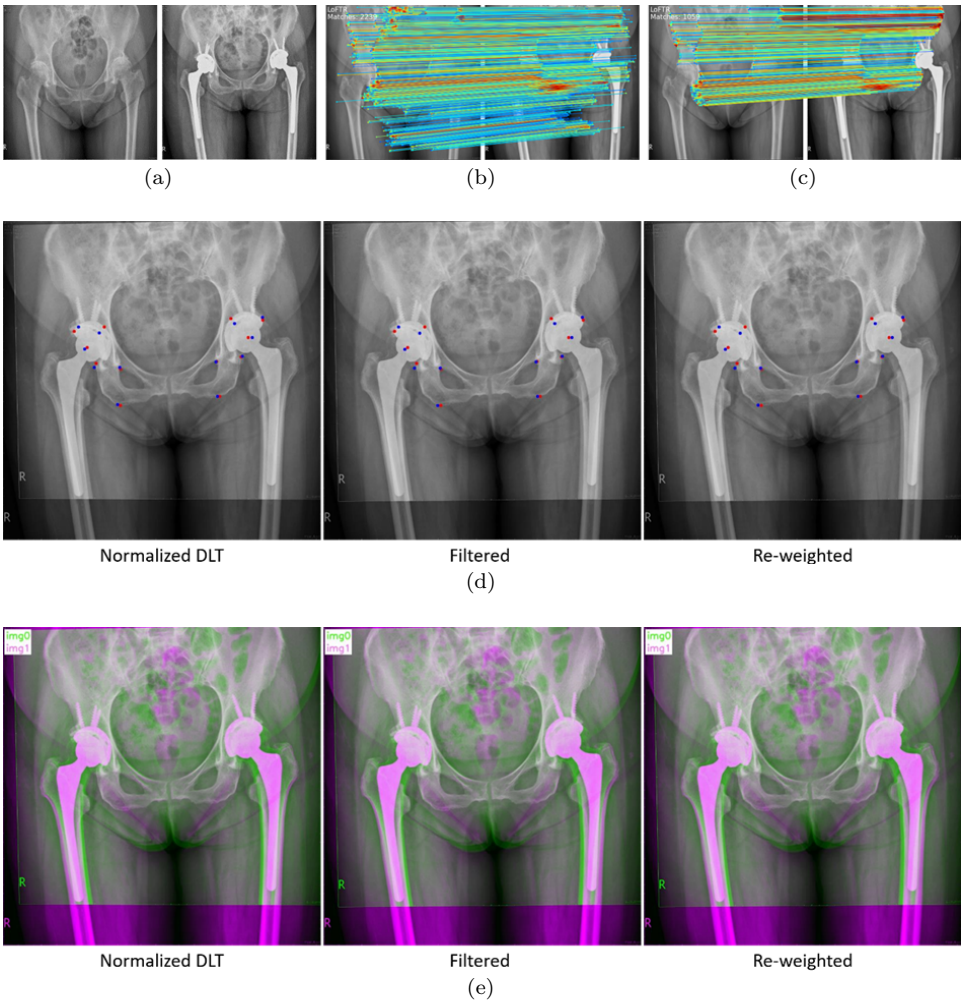


Fig. 14. Alignment result 4. (a) Original image pair. (b) Matching result. (c) Matches within pelvis area. (d) Alignment results with landmarks. (e) CDM. (The MRE are 0.011314, 0.011278, and 0.011282, respectively.)[8]

5. Conclusion

We propose the LinAlign algorithm, which combines the results of semantic segmentation with homography estimation by weights assignment. LinAlign can successfully align the images of non-rigid bodies with partial differences by applying stricter alignment to our interested part. Tested on our dataset, our method has better performance compared with other alignment algorithms.

We also visualize the alignment results in CDM, which can obviously show the difference between two images. With this algorithm, surgeons can ensure the

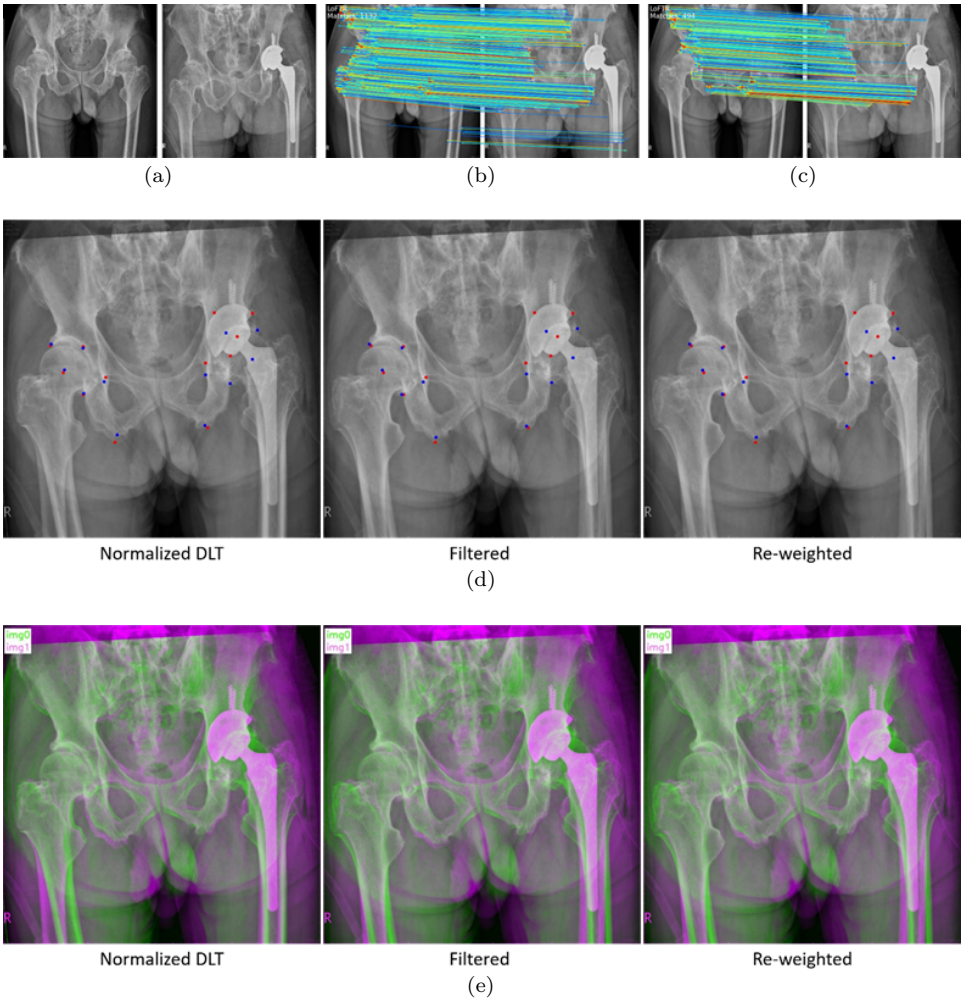


Fig. 15. Failure alignment result. (a) Original image pair. (b) Matching result. (c) Matches within pelvis area. (d) Alignment results with landmarks. (e) CDM. (The MRE are 0.034356, 0.032245, and 0.03253, respectively.)[8]

location of implants is correct during the surgery and can prevent risks such as dislocation after the surgery.

References

1. V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(12), 2481–2495, 2017. DOI: 10.1109/TPAMI.2016.2644615.

2. F. L. Carty, J. P. Cashman, J. Parvizi, A. Zoga, and W. B. Morrison. Imaging of the postoperative hip. *Seminars in Musculoskeletal Radiology*, **15**(4), 357–371, 2011. DOI: 10.1055/s-0031-1286016.
3. L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In *Proceeding of International Conference on Learning Representations*, San Diego, CA, 2015, pp. 1–14.
4. L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of European Conference on Computer Vision, Munich*, Germany, 2018, pp. 833–851.
5. D. Daniel, T. Malisiewicz, and A. Rabinovich. SuperPoint: Self-supervised interest point detection and description. In *Proceedings of CVPR Workshop*, Salt Lake City, Utah, 2018, pp. 337–349.
6. M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, **24**(6), 381–395, 1981.
7. Z. Li and N. Snavely. MegaDepth: Learning single-view depth prediction from internet photos. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 2041–2050. DOI: 10.1109/CVPR.2018.00218.
8. C. W. Lin. LinAlign: X-ray image alignment before and after total hip arthroplasty. Master Thesis, Department of Computer Science and Information Engineering, National Taiwan University, 2023.
9. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.
10. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, **60**(2), 91–110, 2004.
11. D. Nguyen. Simulation and experimental study on polishing of spherical steel by non-Newtonian fluids. *The International Journal of Advanced Manufacturing Technology*, **107**, 763–773, 2020. DOI: 10.1007/s00170-020-05055-w.
12. D. Petek, D. Hannouche, and D. Suva. Osteonecrosis of the femoral head: Pathophysiology and current concepts of treatment. *EFORT Open Reviews*, **4**(3), 85–97, 2019. DOI: 10.1302/2058-5241.4.180036.
13. O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, 2015, pp. 234–241.
14. P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 2020, pp. 4937–4946. DOI: 10.1109/CVPR42600.2020.00499.
15. J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou. LoFTR: Detector-free local feature matching with transformers. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Nashville, TN, 2021, pp. 8918–8927. DOI: 10.1109/CVPR46437.2021.00881.
16. H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia. Pyramid scene parsing network. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 6230–6239, DOI: 10.1109/CVPR.2017.660.
17. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, **39**(6), 1856–1867, 2020. DOI: 10.1109/TMI.2019.2959609.