



Misshapen Pelvis Landmark Detection by Spatial Local Correlation Mining for Diagnosing Developmental Dysplasia of the Hip

Chuanbin Liu¹, Hongtao Xie^{1(✉)}, Sicheng Zhang², Jingyuan Xu¹, Jun Sun²,
and Yongdong Zhang¹

¹ School of Information Science and Technology,
University of Science and Technology of China, Hefei 230026, China
htxie@ustc.edu.cn

² Anhui Provincial Children's Hospital, Hefei 230026, China

Abstract. Developmental dysplasia of the hip (DDH) refers to an abnormal development of the hip joint in infants. Accurately detecting and identifying the pelvis landmarks is a crucial step in the diagnosis of DDH. Due to the temporal diversity and pathological deformity, it is a difficult task to detect the misshapen landmark and diagnose the DDH illness condition for both human expert and computer. Moreover, there is no adequate and public dataset of DDH for research. In this paper, we investigate the spatial local correlation with convolutional neural network (CNN) for misshapen landmark detection. First, we convert the detection of a landmark to the detection of the landmark's local neighborhood patch, which yields effective spatial local correlation for the identification of a landmark. Then, a deep learning based method named FR-DDH network, is proposed for misshapen pelvis landmark detection. It mines the spatial local correlation and detects the best-matched region according to the spatial local correlation. To the end, the landmarks are located at the center of the regions. Besides, a dataset with 9813 pelvis X-ray images is constructed for research in this area, and it will be released for public research. To the best of our knowledge, this is the first attempt to apply deep learning in the diagnosis of DDH. Experimental results show that our approach achieves an excellent precision in landmark location (MAE 1.24 mm) and illness diagnosis over human experts.

Keywords: Developmental dysplasia of the hip · Landmark detection · Spatial local correlation

1 Introduction

Developmental dysplasia of the hip (DDH) refers to a spectrum of hip joint abnormalities ranging from mild acetabular dysplasia to irreducible hip joint dislocation. It is the most common pediatric hip disorder, affecting 0.16% to

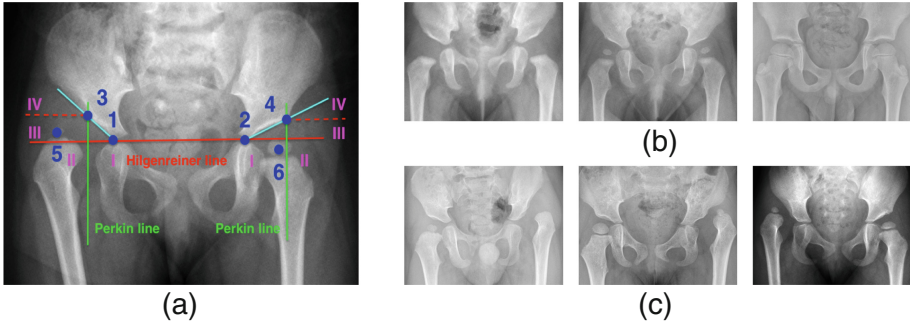


Fig. 1. The diagnosis of developmental dysplasia of the hip. (a) shows the principle of the diagnostic standard, and the key landmarks are: (1) right tri-radiate cartilage center (RTCC), (2) left tri-radiate cartilage center (LTCC), (3) right acetabulum superolateral margin (RASM), (4) left acetabulum superolateral margin (LASM), (5) right femoral head (RFH), (6) left femoral head (LFH). (b) shows the examples of the temporal diversity of DDH. (c) shows the examples of the pathological deformity of DDH.

2.85% of all newborns [1]. The traditional diagnostic methods rely mainly on Xray or Ultrasound images of the pelvis and hip [2,3], and Xray is the primary tool in diagnosing DDH after 6 months of age. Figure 1(a) gives the principle of Xray diagnosis standards. The most important references for Xray DDH diagnosis are the Hilgenreiner’s line, the Perkin’s line and the femoral head, which are strictly relying on the location of pelvis landmarks. However, the landmarks detection for DDH is a challenging task, because (1) during the different stages of skeleton calcification, the landmarks appear with diversity in shape as Fig. 1(b), (2) different grades of dislocation will lead to varying deformity as Fig. 1(c). The temporal diversity and pathological deformity lead DDH diagnosis a time-consuming and experience-sensitive task for orthopedists. Therefore, it suffers from high inter-exam variability and low accuracy. With the development of machine learning [4,5], to overcome these defects, a series of Computer-Aided Diagnosis (CAD) methods have been proposed [1,6–9].

Related Work: Several CAD methods have been proposed for Xray DDH diagnosis. Bashir et al. [8] propose an edge detection method to measure the acetabular angle from the X-ray images. But the miscalculating “*can result from the incomplete development of femur head for infants less than 6 months*”. Similarly, Sahin et al. [9] present a template-matching method for measuring acetabular angles by finding the obturator foramen. However, patients with “*distorted shape of the obturator foramen are not suitable for this approach*”. Bier et al. [10] put forward a sequential prediction framework to detect pelvic anatomical landmarks. Yet, it exhibits poor robustness and “*is susceptible to scenarios not included in training*”. To sum up, existing methods are inapplicable to deal with the temporal diversity and pathological deformity in DDH.

Recently, Arik et al. [11] propose a convolutional neural network system for cephalometric landmarks detection. To overcome the deformity of pathological cases, an image patch with pre-defined size centered at landmark l is extracted as the local neighborhood. The local neighborhood yields effective spatial local correlation for the identification of a landmark, and CNN exhibits well-suited performance in exploiting spatial local correlation by imposing local connectivity patterns. This method performs a CNN forward pass on each sliding window without sharing computation. Consequently, the training is expensive in space and time, and the landmark detection is slow.

Contribution: The local neighborhood around a landmark yields effective spatial local correlation, which can be strong identification of the landmark. To overcome the temporal diversity and pathological deformity challenge in DDH, in this paper, we convert the detection of a landmark to the detection of the landmark’s local neighborhood patch. Then, a deep learning based method named FR-DDH network, is proposed for pelvis landmark detection. It mines the spatial local correlation and detects the best-matched region with CNN. To the end, the landmarks are located at the center of the regions. Besides, a dataset with 9813 pelvis X-ray images is constructed for research in this area, which will be public in the future. To the best of our knowledge, this is the first attempt to apply deep learning in the diagnosis of DDH. Experimental results show that our approach achieves a excellent precision in landmark location (MAE 1.24 mm) and illness diagnosis over human experts.

2 Method

Overall Framework: Figure 2 illustrates the overall FR-DDH framework for Xray DDH diagnosis. The neighbourhood image patch centered at landmark l is extracted as detection target, and FR-DDH is trained to detect the patch from a pelvis image. For an input image, a series of convolutional layers are applied to mine the spatial local correlation and generate the high-dimensional feature map. Then the local neighborhood region proposals are generated by Region Proposal Network (RPN), according to the feature map. Combing the region proposals and the feature map by ROI pooling, FR-DDH predicts the categories of the region and their bounding-box regression offsets, and generates the detection result of each image patch. Finally, the specific landmark is located at the center of the patch, and we get the diagnosis result according to the landmarks.

Local Image Patch Extraction: To detect the landmark l with temporal diversity and pathological deformity, the spatial local correlation around landmark l should be learned from the images in the training set. We extract the $(2N + 1) \times (2N + 1)$ image patch centered at landmark l as the local neighborhood, as Fig. 2 shows, where N is sufficiently large to visually recognize the landmark. Hence, we convert the detection of a landmark to the detection of

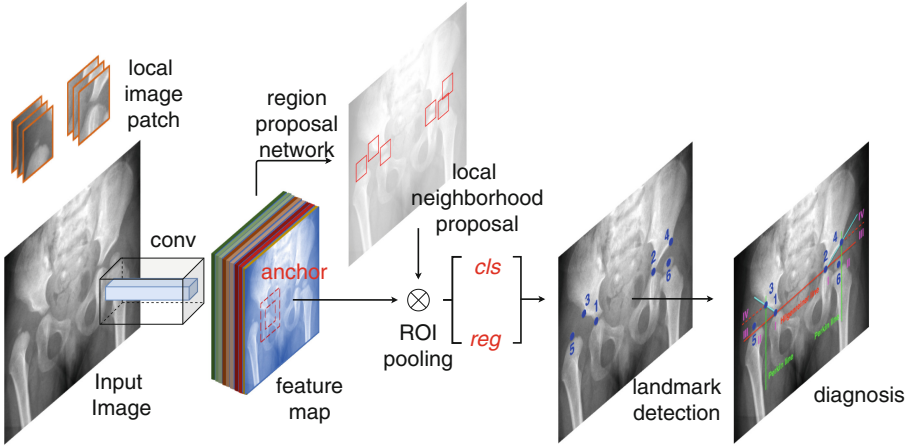


Fig. 2. The framework of our FR-DDH. [Best viewed in color] (Color figure online)

the landmark’s local neighbourhood patch, which yields effective spatial local correlation for the identification of a landmark.

Spatial Local Correlation Mining: In FR-DDH, We use ResNet101 with weights trained on ImageNet as feature extraction network. ResNet101 exhibits strong ability in mining spatial local correlation by imposing local connectivity patterns and merging feature map with skip connection. The images are rescaled to $h \times 600 \times 3$ by repeating 3 times to use pretrained weights. The shorter side is rescaled to 600 while the longer side is rescaled to h . After a series of hierarchical *conv*, FR-DDH mines the spatial local correlation and outputs a 2048-D feature map.

Region Proposal and Landmark Detection: Figure 3 illustrates the framework of region proposal and landmark detection of FR-DDH. RPN uses the generated 2048-D feature maps for generating local neighborhood region proposals, each with an objectness score. As proposed by Faster-RCNN [12], we slide a network over the convolutional feature map of the conv5 layer in a sliding-window fashion. This network is fully connected to a spatial window of the convolutional feature map with a 3×3 convolutional layer. Region proposals are relative reference boxes to anchors centered at each sliding window. Each anchor is related with a scale of size 128 and 256 pixels and aspect ratios of 1 : 1.

Once the local neighborhood region proposal is generated, FR-DDH combines the region proposals and the feature map by ROI pooling. Each proposal is pooled into a fixed-size feature map and then mapped to a feature vector by fully connected layers. Then, each feature vector branches into two sibling output layers: *cls* layer for classifying the categories of local neighborhood, and *reg* layer

for regressing the bounding box coordinates. The landmark is finally detected on the center of the local neighborhood region.

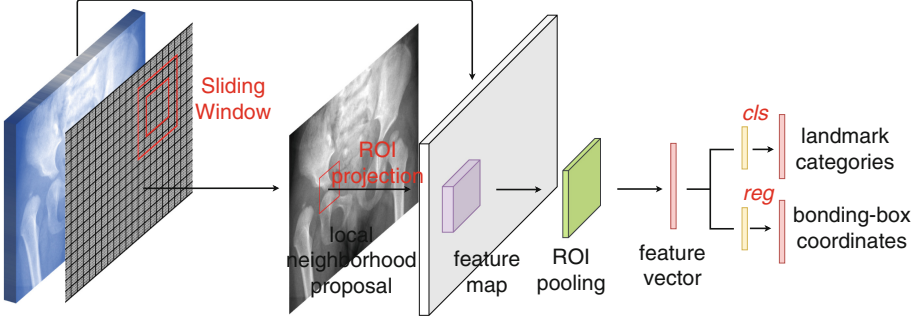


Fig. 3. The framework of Region proposal and landmark detection.

Loss Function for Learning: We minimize an objective function following the multi-task loss in Faster R-CNN [12]. Our loss function for an image is defined as:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

The classification layer cls outputs a discrete probability $\{p_i\}$ ($0 \leq i \leq K$) over $K+1$ (landmarks + background) categories and the regression layer reg outputs $\{t_i\}$ bounding-box regression offsets a predicted tuple $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$ for class u . Here, i is the index of an anchor in a mini-batch and p_i is the predicted probability of anchor i being an local neighborhood patch. The ground-truth label p_i^* is 1 if the anchor is positive, and is 0 if the anchor is negative. t_i is a vector representing the 4 parameterized coordinates of the predicted bounding box, and t_i^* is ground-truth box associated with a positive anchor.

The classification loss L_{cls} is a log loss over $K+1$ categories:

$$L_{cls}(p, u) = -\log p_u. \quad (2)$$

The regression loss L_{reg} is a smooth L1 function:

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (3)$$

The term $p_i^* L_{reg}$ means the regression loss is activated only for positive anchors ($p_i^* = 1$) and is disabled otherwise ($p_i^* = 0$). The two terms are normalized with N_{cls} , N_{reg} and a balancing weight λ , which is set to 10.

3 Experiments and Results

Data: We note that there is no public DDH dataset, which seriously limits the research on diagnosing DDH. To employ deep learning in the diagnosis of DDH, a dataset with adequate pelvis images is required. Accordingly, in this paper, 24000 X-ray images of pelvis are collected and resampled with pixel spacing as 0.15 mm. After the strict screening from the orthopedist, 9813 images of them are kept in the dataset, with 7710 for training and 2103 for testing. The age of each case ranges from 3 months to 12 years, and the illness involves normal to terrible dislocation. To the best of our knowledge, this is the first dataset for DDH and the dataset will be public for research¹.

Experiment Setup: Our FR-DDH is implemented with PyTorch, an optimized tensor library for deep learning. We randomly initialize all new layers by drawing weights from a zero-mean Gaussian distribution with standard deviation of 0.01. And the other layers (i.e., the shared convolutional layers) are initialized by ResNet101 pretrained from ImageNet. We use a learning rate of 0.001 for 80k mini-batches, and 0.0001 for the next 30k mini-batches on the dataset. The momentum is set to be 0.9 and the weight decay is set to be 0.0005. The FR-DDH is trained on a Ubuntu workstation with one NVIDIA GeForce 1080Ti GPU, and it takes one day for training the model.

Evaluation Metric: To validate the accuracy of our method, we define the landmark-specific point-to-point error for landmark l as

$$PEL_l = \left(\sum_{i=1}^n \|m_{li} - a_{li}\| \right) / n. \quad (4)$$

Here n represents the number of images, m represents the manually labeled landmarks and a represents the automatically identified landmark. The average point-to-point errors (PE) is defined as the average of PEL_l as

$$PE = \sum_{l=1}^k \frac{PEL_l}{k} = \sum_{l=1}^k \sum_{i=1}^n \frac{\|m_{li} - a_{li}\|}{nk}. \quad (5)$$

Here k represents the number of landmarks. We also report the successful detection rate (SDR) which gives the percentage of images for which a landmark l is located within a precision range $z \in \{1.5 \text{ mm}, 2.0 \text{ mm}, 3.0 \text{ mm}\}$ as

$$SDR_l = \# \{i : \|m_{li} - a_{li}\| \leq z\} / n \times 100 \quad (6)$$

¹ The dataset, diagnosis method and evaluation code will be released at <https://github.com/liuboss1992/FR-DDH>.

Result: A series of experiments have been conducted with different scales of local neighborhood patch, where N ranges from 50 to 100. Table 1 shows the relationship between neighborhood region scale N and average point-to-point error PE . In FR-DDH, the $(2N+1) \times (2N+1)$ image patch centered at landmark l is extracted as the local neighborhood. An image patch with small N may not provide adequate spatial local correlation, hence the detection accuracy will be low. Meanwhile, an image patch with oversized N may introduce extraneous information, which will also lead to low accuracy. We achieve the best accuracy with $PE = 1.244$ mm when $N = 80$.

Table 1. Relationship between neighborhood region scale and point-to-point error. For RTCC, LTCC, ..., LFH, please refer to Fig. 1.

| PE_{l_i} (pixel) | FR-DDH N = 50 | FR-DDH N = 60 | FR-DDH N = 70 | FR-DDH N = 80 | FR-DDH N = 90 | FR-DDH N = 100 |
|--------------------|------------------|------------------|------------------|------------------|------------------|-------------------|
| Average | 1.295 | 1.281 | 1.277 | 1.244 | 1.276 | 1.282 |
| RTCC | 1.464 | 1.518 | 1.449 | 1.459 | 1.464 | 1.459 |
| LTCC | 1.569 | 1.627 | 1.588 | 1.601 | 1.544 | 1.587 |
| RASM | 1.170 | 1.173 | 1.104 | 1.152 | 1.172 | 1.226 |
| LASM | 1.277 | 1.294 | 1.458 | 1.258 | 1.331 | 1.332 |
| RFM | 1.069 | 0.999 | 0.953 | 0.992 | 1.013 | 1.003 |
| LFH | 1.223 | 1.074 | 1.111 | 1.002 | 1.130 | 1.085 |

As is illustrated in Table 2, we conduct contrast experiment with other baseline for measuring the Acetabular Index. We follow Bashir’s work [8] to take absolute error (AE) and average accuracy (AA) as the evaluation metric. Compared with Bashir’s work which employs an edge detection approach for landmark detection, our FR-DDH achieves lower error and higher accuracy. In addition, Bashir evaluates its model on only 24 infants. By contrast, our FR-DDH is evaluated on a wider variety of 2000+ infants. The comparison fully shows the reliability and robustness of FR-DDH.

Table 2. Performance on measuring the Acetabular Index.

| Method | AE | | AA | |
|------------|--------------|--------------|---------------|---------------|
| | Right | Left | Right | Left |
| Bashir [8] | 3.78° | 2.95° | 83.60% | 85.40% |
| FR-DDH | 2.32° | 2.36° | 87.80% | 87.20% |

Figure 4 presents the successful detection rate for each landmark of FR-DDH, when $N = 80$. Almost 95% landmarks can be detected within $z = 3$ mm, which

is a reliable performance for clinical use. With the accurate detection of landmarks, FR-DDH further diagnoses the illness of DDH. Compared with the diagnosis result from domain expert, FR-DDH obtains precision of 92.8% and recall of 97.5%. By contrast, a general doctor obtains precision of 89.9% and recall of 91.5% in our research. FR-DDH achieves excellent performance in illness diagnosis over human experts. The details of the diagnosis code will be released in our provided link.

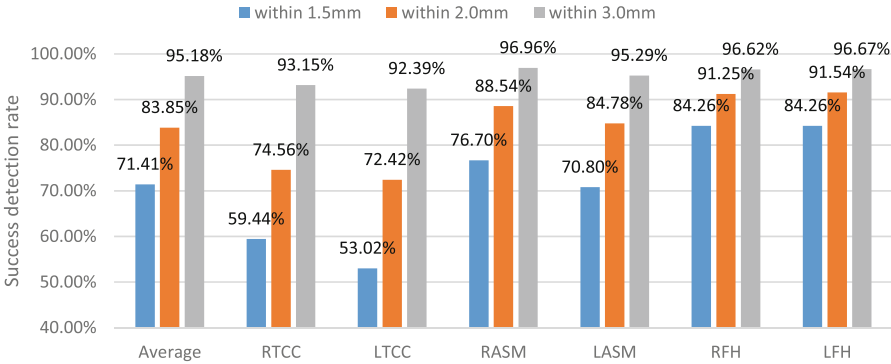


Fig. 4. Success detection rates of each landmark in FR-DDH, when N is set to be 80.

4 Conclusion

This paper puts forward FR-DDH, a novel approach for misshapen pelvis landmarks detection of DDH by mining the spatial local correlation of neighborhood region. The temporal diversity and pathological deformity bring challenges for anatomical landmark detection. We investigate the spatial local correlation for misshapen landmark detection, and convert the detection of a landmark to the detection of the landmark's local neighborhood patch. Besides, a dataset with 9813 pelvis X-ray images is constructed for this task, and it will be released for public research. This work can be an enlightening reference and be generalized for numerous anatomical landmark detection tasks.

Acknowledgements. This work is supported by the Huawei-USTC Joint Innovation Project on Machine Vision Technology (FA2018111122).

References

1. Quader, N., Hodgson, A.J., Mulpuri, K., Cooper, A., Abugharbieh, R.: A 3D femoral head coverage metric for enhanced reliability in diagnosing hip dysplasia. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 100–107. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_12

2. Ruiz Santiago, F., et al.: Imaging of hip pain: from radiography to cross-sectional imaging techniques. *Radiol. Res. Pract.* **2016** (2016)
3. Atweh, L.A., Kan, J.H.: Multimodality imaging of developmental dysplasia of the hip. *Pediatr. Radiol.* **43**(1), 166–171 (2013)
4. Liu, A.A., Su, Y.T., Nie, W.Z., Kankanhalli, M.: Hierarchical clustering multi-task learning for joint human action grouping and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(1), 102–114 (2016)
5. Xie, H., Yang, D., Sun, N., Chen, Z., Zhang, Y.: Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recogn.* **85**, 109–119 (2019)
6. Paserin, O., Mulpuri, K., Cooper, A., Hodgson, A.J., Garbi, R.: Real time RNN based 3D ultrasound scan adequacy for developmental dysplasia of the hip. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11070, pp. 365–373. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_42
7. Paserin, O., Mulpuri, K., Cooper, A., Hodgson, A.J., Abugharbieh, R.: Automatic near real-time evaluation of 3D ultrasound scan adequacy for developmental dysplasia of the hip. In: Cardoso, M.J., et al. (eds.) *CARE/CLIP -2017*. LNCS, vol. 10550, pp. 124–132. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67543-5_12
8. Al-Bashir, A.K., Al-Abed, M., Sharkh, F.M.A., Kordeya, M.N., Rousan, F.M.: Algorithm for automatic angles measurement and screening for developmental Dysplasia of the Hip (DDH). In: 37th Annual International Conference of the IEEE, EMBC 2015, pp. 6386–6389. IEEE (2015)
9. Sahin, S., Akata, E., Sahin, O., Tuncay, C., Özkan, H.: A novel computer-based method for measuring the acetabular angle on hip radiographs. *Acta orthopaedica et traumatologica turcica* **51**(2), 155–159 (2017)
10. Bier, B., et al.: X-ray-transform invariant anatomical landmark detection for Pelvic Trauma surgery. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) *MICCAI 2018*. LNCS, vol. 11073, pp. 55–63. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_7
11. Arik, S.Ö., Ibragimov, B., Xing, L.: Fully automated quantitative cephalometry using convolutional neural networks. *J. Med. Imaging* **4**(1), 014501 (2017)
12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems*, pp. 91–99 (2015)