2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

XAI Feature Detector for Ultrasound Feature Matching

Zihao Wang, Hang Zhu, Yingnan Ma, Anup Basu
Department of Computing Science, University of Alberta

Abstract— **Feature matching is a crucial component of computer vision that has various applications. With the emergence of Computer-Aided Diagnosis (CAD), the need for feature matching has also emerged in the medical imaging field. In this paper, we proposed a novel algorithm using the Explainable Artificial Intelligence (XAI) [1] approach to achieve feature detection for ultrasound images based on the Deep Unfolding Super-resolution Network (USRNET). Based on the experimental results, our method shows higher interpretability and robustness than existing traditional feature extraction and matching algorithms. The proposed method provides a new insight for medical image processing, and may achieve better performance in the future with advancements of deep neural networks.**

I. Introduction

Ultrasonography is a widely used methodology for the early clinical detection of various diseases. It plays a pivotal part in clinical image based diagnosis for its efficiency, low-cost, and convenience. For diagnostic accuracy, feature matching is widely used in medical image processing since the results can provide decision support for medical diagnosis. Consequently, computer aided diagnosis systems are required in medical institutions to assist doctors in analyzing and interpreting ultrasound images to improve the accuracy, objectivity, and repeatability of examinations. Our task has two components: feature detection and feature matching. Feature detection segments and localizes significant interesting points from the original ultrasound images. Feature matching is used to visualize the correspondences between features. Ultrasound images are different from regular images since they usually suffer from severe noise, poor image contrast, and shadowing artifacts, which makes it difficult for doctors to identify whether patches of images are similar or not.

In the state-of-the-art, many traditional methods for feature detection has been proposed. Scale Invariant Feature Transform (SIFT) [2] is a well-known detector, which utilizes scale-invariant information and can adapt to rotation and illumination changes. It can also avoid affine transformation to some extent. Speeded Up Robust Feature (SURF) [3] is an optimized version of SIFT. SURF constructs Hessian Matrix to get all potential interesting points for final extraction. Furthermore, SURF decreases feature descriptors to achieve speedup. Differing from SIFT and SURF, Features from Accelerated Segment Test (FAST) [4] applies corner detection while setting a threshold and using non-maximum suppression to reach higher accuracy. Oriented FAST and Rotated BRIEF (ORB) [5] is based on the foundation of FAST. For optimization, ORB applies BRIEF [6] as the descriptor to save computational resources. However, these algorithms are mainly focused on synthetic images and do not have promising feature matching results on ultrasound images.

To represent an algorithm with high interpretability and robustness, we creatively utilize the XAI approach to achieve feature detection and extraction on ultrasound images. Specifically, we implement guided back-propagation to detect and extract features. The guided back-propagation technique was originally used for the interpretability and optimization of classification networks. We explore this XAI approach to generate feature maps based on gradient information. To capture the gradient information, we implement guided back-propagation on state-of-the-art super-resolution neural networks. In particular, we utilize the Deep Unfolding Super-Resolution Network (USRNET) [7] to generate gradients. Guided back-propagation was utilized to achieve visualization by generating feature maps based on these gradient information. To test the performance of our proposed method, we collected data using our handheld ultrasound device.

II. Methodology

In this section, we separate the proposed method into two subsections. The first subsection introduces the proposed XAI feature detector, which creatively utilizes guided back-propagation [8] to detect features based on USRNET [7]. The second subsection introduces the Robust Independent Elementary Features (BRIEF) [6] descriptor and Brute-force feature matcher based on the detected features. The overall architecture of the proposed XAI feature detector is shown in Fig.1.

A. Feature detection

1) Guided Back-propagation: The concept of Explainable Artificial Intelligence (XAI) is usually used by researchers to create a clear insight of deep neural networks. Guided back-propagation is a popular XAI technique, which was originally utilized in classification networks. The original guided back-propagation algorithm can generate saliency maps to reveal the most significant features that impact the classification tasks. In our proposed XAI feature detector, we explore the guided back-propagation as a feature detector because of its feature tracking speciality. However, differing from the original application of guided back-propagation with classification networks, we apply guided back-propagation on super-resolution neural networks. The super-resolution neural network can effectively enhance the image quality

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

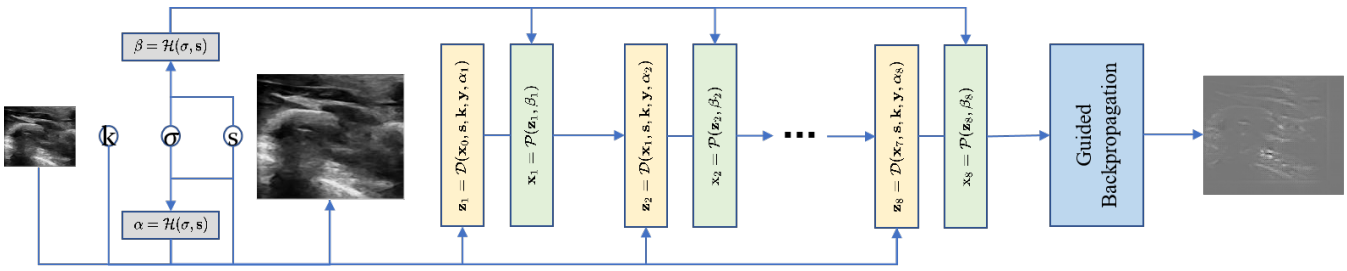

Fig. 1. The architecture of the proposed XAI feature detector.

of high-frequency regions in images. By utilizing a super-resolution network, guided back-propagation can reveal the features that have the greatest impact on the super-resolution network. In this way, we can obtain the most significant high-frequency feature information. Specifically, we implement the Deep Unfolding Super-resolution Network (USRNET), which is introduced in Sub-section 2.

Guided back-propagation [8] was proposed based on vanilla back-propagation [9] and the deconvnet algorithm [10]. The computations of vanilla back-propagation, DeconvNets and guided back-propagation can be summarized in Fig.2. These algorithms are essentially identical except for the computations when passing through the ReLU function of convolution networks. The equation of ReLU activation function (1) is:

$$f_i^{l+1} = relu\left(f_i^l\right) = \max\left(f_i^l, 0\right) \tag{1}$$

In the forward pass, ReLU functions are handled by zeroing out the negative gradient values. In vanilla back-propagation, values that are zeroed out in the forward pass will also be zeroed out during back-propagation. The back-propagation equation (2) is shown below, where f represents the feature maps generated by layers. R is an intermediate calculation result of the back-propagation.

$$R_i^l = \left(f_i^l > 0\right) \cdot R_i^{l+1}, \text{ where } R_i^{l+1} = \frac{\partial f^{out}}{\partial f_i^{l+1}} \tag{2}$$

In the deconvnet algorithm, the gradients travel back to the image space through a deconvnet, which contains operations like unpooling and deconvolution. In deconv implementation, negative values are set to zero. Equation (3) summarizes this step below.

$$R_i^l = \left(R_i^{l+1} > 0\right) \cdot R_i^{l+1} \tag{3}$$

Guided Back-propagation basically acts as a combination of vanilla back-propagation and DeconvNets when passing through the ReLU non-linearity. Blocks with negative values, either in the forward or in backward pass, are set to zero. More precisely, similar to deconvnet, guided back-propagation only allows positive error signals to travel back, also like vanilla back-propagation, guided back-propagation limits the input to positive values, as shown in Equation (4).

$$R_i^l = \left(f_i^l > 0\right) \cdot \left(R_i^{l+1} > 0\right) \cdot R_i^{l+1} \tag{4}$$

The computations of vanilla back-propagation, DeconvNets and guided back-propagation can be summarized in Fig.2. To generate the most significant feature maps with the least independent noise, we implement guided back-propagation on USRNET.

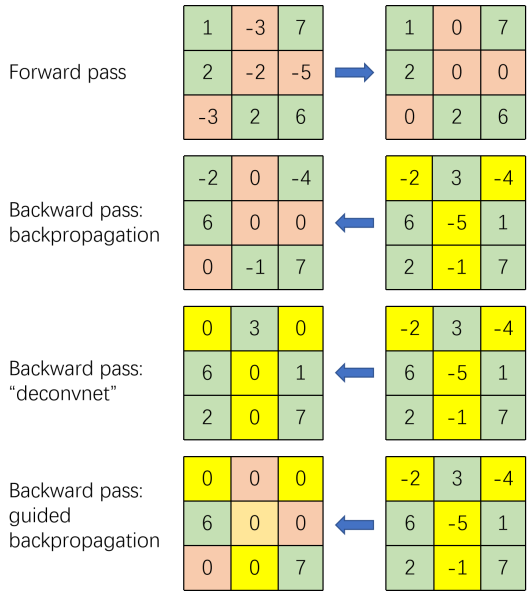

Fig. 2. Calculation maps of forward pass, vanilla back-propagation, deconvnet and guided back-propagation.

2) Unfolding Super-resolution Network: To achieve feature detection, we apply guided back-propagation on [7] the deep unfolding super-resolution network (USRNET). The architecture of USRNET is shown in Fig.1. USRNET combines features from both learning-based and model-based methods for super resolution tasks [7]. It focuses on giving the flexibility of handling different scale factors, blur kernels and noise levels under a unified Maximum A Posteriori (MAP) [11] framework in a single image super-resolution scenario. The network consists of three major modules: data module, prior module, and hyper-parameter module.

Data module $\mathcal{D}$: This module constructs an output image of higher resolution $\mathbf{z}_k$ from the input of original image $\mathbf{y}$, scale factor $\mathbf{s}$, output of the previous prior module $\mathbf{x}_{k-1}$, and the trade-off hyper-parameter α_k. The process can be summarized as minimizing a weighted combination of the

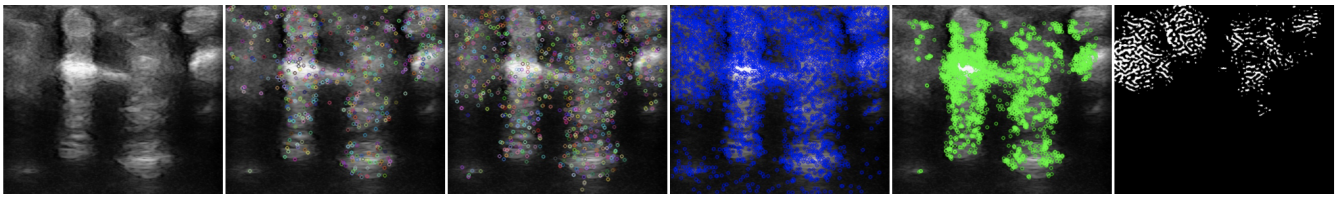

Fig. 3. Comparison of features extracted by different algorithms. The first image is the original. The second image reveals the feature points detected by SIFT. The third image is processed by SURF. The fourth and fifth images are processed by FAST and ORB respectively. The last image is the feature map generated by our proposed method.

data term $\left\|\mathbf{y}-(\mathbf{z}\otimes\mathbf{k})\downarrow_s\right\|$ and the quadratic regularization term $\left\|\mathbf{z}-\mathbf{x}_{k-1}\right\|^2$. The module [7] can be denoted as below.

$$\mathbf{z}_k = \mathcal{D}(\mathbf{x}_{k-1}, \mathbf{s}, \mathbf{k}, \mathbf{y}, \alpha_k) \tag{5}$$

Prior module $\mathcal{P}$: Following the data module, a prior module is applied to denoise the processed image $\mathbf{z}_k$. The denoising procedure is managed by a ResUNet [12]; thus inheriting the ability of fast training and large capacity with residual blocks. The network takes $\mathbf{z}_k$ from the data module and a noise level map β_k. The output is the denoised image $\mathbf{x}_k$, which can be used as the data module input image for the next iteration. The prior module can be described as below.

$$\mathbf{x}_k = \mathcal{P}(\mathbf{z}_k, \beta_k) \tag{6}$$

Hyper-parameter module $\mathcal{H}$: A hyper-parameter module is employed to control the output of the data and prior modules. Specifically, the module predicts sets of α and β from scale factor $\mathbf{s}$ and noise level σ, combined with a constant trade-off parameter λ and penalty parameter μ. With proper set-up of these modules, the network can achieve the task of unfolding optimization; thus recovering higher resolutions of the input image. More precisely, compared to input images, those output counterparts not only have higher resolution, but also details and edges. In other words, most significant features are also enhanced by the super-resolution neural network. Guided back propagation can generate feature maps that contain the gradients of various areas.

B. Feature Matching

After obtaining the feature gradient maps, a thresholding method is applied to the map to binarize the points in the map. Thus, the point locations can be conveniently extracted and represented in a coordinate format. Then, the Robust Independent Elementary Features (BRIEF) descriptor [6] is employed to give identities to each point. BRIEF first performs a Gaussian blur on the entire image to reduce noise. For each given feature point, it randomly selects pairs of locations within a square patch centered at a point. Each pair of locations gives a binary result based on the following equation, where $\mathbf{p}$ is the patch and $\mathbf{x}$, $\mathbf{y}$ is a pair of locations.

$$\tau(\mathbf{p};\mathbf{x},\mathbf{y}) := \begin{cases} 1 & \text{if } \mathbf{p}(\mathbf{x}) < \mathbf{p}(\mathbf{y}) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

With the results from all the pairs, a binary string can be constructed as:

$$f_{n_d}(\mathbf{p}) := \sum_{1\le i\le n_d} 2^{i-1}\tau\left(\mathbf{p};\mathbf{x}_i,\mathbf{y}_i\right) \tag{8}$$

Finally, the descriptors and their corresponding points from two images are matched through a brute-force matching method [13]. The L2 norm is calculated over the Cartesian products of the two set of interest points. Only the match with highest L2 norm value is considered as a match. Then, matches are pruned using the Random Sample Consensus (RANSAC) algorithm [14] to iteratively eliminate low quality matches.

III. Experiments

In this section, we present the detection and matching results of the proposed XAI detector. We also compare with other state-of-the-art methods to demonstrate improvement.

A. Data

The data was collected using the Clarius handheld ultrasound device focusing on two parts of the human body; namely, knuckles and heart. To ensure consistency of feature positions and the time differences between the target and reference images, we capture adjacent frames of an ultrasound video as the experimental data. Consequently, when implementing feature matching, straight lines represent correct matches and oblique lines represents wrong matches.

B. Feature detection

First, we compare the feature detection results of the proposed XAI Feature detector with state-of-the-art feature detectors; namely, SIFT, SURF, FAST and ORB. The feature detection results are shown in Fig.3. As the figure illustrates, the feature points detected by SIFT and SURF are sparsely distributed instead of being in a concentrated region. When it comes to the FAST detector, the extracted feature points mainly reveal the significant knuckle features. However, FAST detector also detects some insignificant noisy points at the bottom of the demo image. Compared to FAST, the ORB detector can detect feature points concentrating in high-frequency regions. But ORB cannot successfully detect the features near the image boundaries. Compared to state-of-the-art detectors, the proposed XAI detector can achieve outstanding performance. To be specific, the XAI detector

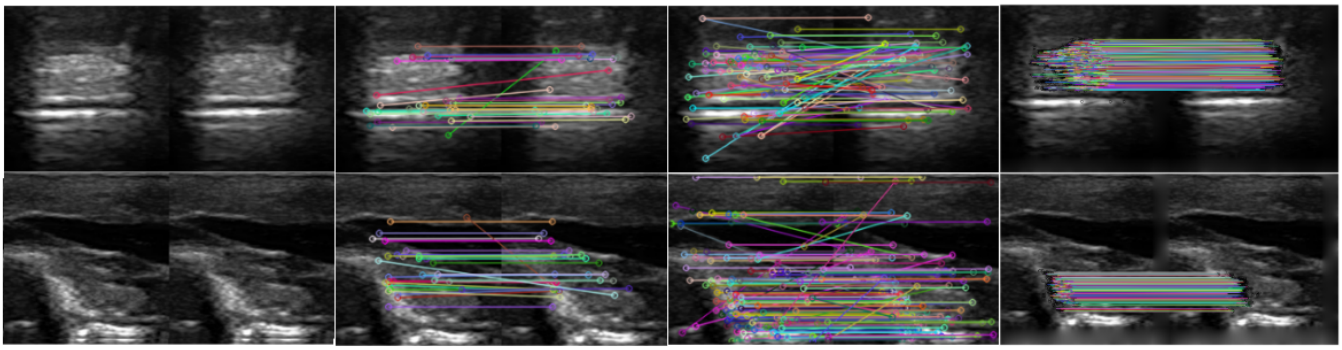

Fig. 4. Feature matching results on sample images. The two rows represent two different pairs of matching images separately. The first column illustrates two pairs of target and reference images. The second column shows the matching results based on SIFT. The third column represents the matching results using ORB detector. The last column shows the matching results using the proposed XAI detector.

can detect feature points localized in high-frequency regions. The detected feature points can reveal the bone texture of target areas without including independent noise points.

C. Feature matching

After the features are detected, matchings are computed using the BRIEF descriptor and a brute-force matcher. The performance is evaluated against ORB and SIFT with identical descriptor and matcher. The matching quality is determined by two important factors: the interest point coverage and the accuracy of the descriptor. In our experiment, the descriptor is uniform. Thus, the major difference is based on the quality and quantity of interest points. Our method outperforms ORB and SIFT by involving the corresponding interest points across images. Fig.4 shows a side by side comparison between the algorithms. From Fig.4 we can see that the matching for the ORB algorithm has more correct matches than SIFT. But, the ORB detector also leads to more incorrect matches due to the detection of independent noise. The SIFT algorithm produces less matches compared with ORB, but the amount of incorrect matches is also decreased. Finally, our method provides large amount of high quality matches over the high frequency area and few wrong matches. The proposed XAI detector outperforms ORB and SIFT in terms of matching accuracy.

IV. Conclusion

In this paper, we proposed an XAI based feature detector, which explores guided back-propagation to extract significant features based on USRNET. The detection and matching results outperform state-of-the-art detector. The XAI detector generates promising results and provides new inspiration for the study of automatic feature extraction and matching. Due to the nature of USRNET, the feature detection was focussed on high frequency areas. In the future, more effort will be put on extending this method to the entire image. We believe that the proposed method has great potential in identifying and linking organ parts between medical images, thereby contributing to the feasibility of automatic medical image processing.

References

[1] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519308103

[2] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 11 2004.

[3] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417.

[4] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European conference on computer vision*. Springer, 2006, pp. 430–443.

[5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International Conference on Computer Vision*, 2011, pp. 2564–2571.

[6] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *European conference on computer vision*. Springer, 2010, pp. 778–792.

[7] K. Zhang, L. Van Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3217–3226.

[8] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

[9] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[10] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[11] R. Bassett and J. Deride, "Maximum a posteriori estimators as a limit of bayes estimators," *Mathematical Programming*, vol. 174, no. 1, pp. 129–144, 2019.

[12] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.

[13] A. Jakubović and J. Velagić, "Image feature matching and object detection using brute-force matchers," in *2018 International Symposium ELMAR*. IEEE, 2018, pp. 83–86.

[14] 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*, vol. 24, no. 6, pp. 381–395, 1981.