Medical Instrument Detection in Ultrasound-Guided Interventions: A Review

Hongxu Yang, Caifeng Shan, Alexander F. Kolen, Peter H. N. de With

***Abstract*—Medical instrument detection is essential for computer-assisted interventions since it would facilitate the surgeons to find the instrument efficiently with a better interpretation, which leads to a better outcome. This article reviews medical instrument detection methods in the ultrasound-guided intervention. First, we present a comprehensive review of instrument detection methodologies, which include traditional non-data-driven methods and data-driven methods. The non-data-driven methods were extensively studied prior to the era of machine learning, i.e. data-driven approaches. We discuss the main clinical applications of medical instrument detection in ultrasound, including anesthesia, biopsy, prostate brachytherapy, and cardiac catheterization, which were validated on clinical datasets. Finally, we selected several principal publications to summarize the key issues and potential research directions for the computer-assisted intervention community.**

***Index Terms*—Medical instrument, ultrasound-guided interventions, B-mode, review.**

I. Introduction

With increasing financial pressure on the healthcare system a general trend toward efficient workflow, shortening procedure time and higher clinical outcome resulting in less re-dos for any given intervention or surgery is preferred. To guide intervention operation, e.g. cardiac intervention, advanced medical imaging systems such as ultrasound and X-ray are required, which offer surgeons accurate visualization and measurement of anatomical structures while displaying the interventional activities within or outside the operation regions. These systems enable the guidance of a medical instrument inside the patient body without open incision by adopting a needle puncture or catheter insertion, which is commonly known as image-guided minimally invasive intervention [1], [2], [3], [4]. This approach is being increasingly adopted in many surgical applications because of its lower risk of complications, shorter patient recovery time and reduced cost. Moreover, with advanced computer vision and signal processing techniques, it is possible to obtain operation-related data with higher quality in an efficient way, which enables researchers to analyze the data and provide solutions to guide the procedure.

The importance of medical imaging systems and the amount medical data are rapidly growing in the past years. These strong developments in medical imaging enable novel applications to be applied in the area of minimally invasive surgery. Among the key imaging modalities used in image-guided minimally invasive surgery, such as fluoroscopy and ultrasound (US), US imaging has received significant attention in recent years, because of its advantages, like wide availability, non-ionizing and real-time performance [1] compared to X-ray, CT or MRI imaging. Furthermore, US offers the unique benefits of a wide range of transducers that can be used in different application scenarios from operation room to emergency medical units [5]. As a consequence, US-guided interventional procedures are broadly investigated and applied in different applications, such as biopsy [6], [7], regional anesthesia [8], ablation therapy [9], [10], prenatal diagnosis and therapy [11] and structural and congenital heart disease. With US guidance during minimally invasive interventions, automatically instrument detection could be beneficial to shorten the procedure time. It would simplify the manipulations of the ultrasound transducer, which reduces the time for the sonographer to identify the instrument inside the body and perform the intervention. Therefore, the total complexity is reduced with a simplified procedure, which is benefit to both patient and interventionalist, especially for the therapies with exposure to radiation, like cardiac catheterization.

Existing approaches for instrument detection can be classified as follows. First, instrument detection based on external or internal sensing devices, such as optical fiber sensing [12], Electromagnetic tracking [13], and robotic-guided detection [14]. Second, purely image-based approaches, without applying any supplementary sensors or devices. Although additional sensing-based methods have achieved promising results, the relatively high cost of equipment and the involved sensors complicated system set up in the operation room have hampered their broad acceptance. In contrast, many image-based approaches have been proposed to detect the medical instrument in US images. Starting with the instrument modeling by simple ultrasound image intensity analysis in the early work up to the latest developed deep learning-based segmentation, various approaches have been introduced. As illustrated in Fig. 1 (a), the first work on medical instrument detection appeared in 2002, though the number of papers grew rapidly after 2012. Based on extensive analysis and literature review, this paper presents a broad review on image-based instrument detection in ultrasound for minimally invasive interventions. To the best of our knowledge, this is the one of the first comprehensive reviews on medical instrument detection in US imaging.

The review article is organized as follows. First, Section II involves the review of the methodologies, which clusters the related literature into non-data-driven and data-driven

Hongxu Yang and Peter H. N. de With are with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

Caifeng Shan is with Shandong University of Science and Technology, Qingdao, China.

Alexander F. Kolen is with Philips Research, Eindhoven, The Netherlands.

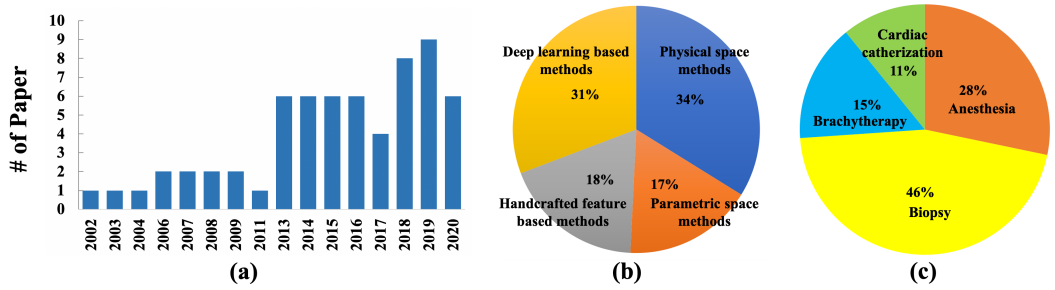

Fig. 1. Breakdown of the papers included in this review (a) in year of publications until April 2020, (b) instrument detection methods (Section II), and (c) major clinical applications (Section III). The total number of papers for this review article is 65, the last update to the included paper was on June 1, 2020.

approaches. More specifically, non-data-driven methods are grouped as they are based on a-priori assumptions on the medical instrument shape via the local image intensity distribution of voxels or pixels. As a consequence, it mainly includes parametrical space-based methods and physical space-based methods, since they detect the instrument in parametric transformation and normal physical space, respectively. Data-driven methods are based on vastly studied machine learning approaches. More specifically, the instrument-related information can be modeled and learned from the data itself, and is categorized into handcrafted feature-based methods and recent deep learning-based methods. The above classification is based on the dominant methodology of the literature, which can also include pre-/post-processing in the considered framework. The overview of the ratio of published papers is summarized in Fig. 1 (b). Second, Section III discusses the related clinical applications, such as anesthesia therapy, biopsy, prostate brachytherapy and cardiac interventions with catheters, of which a literature overview is summarized in Fig. 1 (c). Third, Section IV systematically discusses the evaluation methods, dataset descriptions and experimental performance of these papers. Fourth, Section V presents examples of the principal publications in the area. Meanwhile, it includes discussions on the key issues and potential research directions in the computer-assisted intervention area based on these principal articles. Finally, conclusions are given in Section VI.

II. Instrument Detection Methods

In this section, we have segmented the instrument detection methods into two classes: non-data-driven methods and data-driven methods. Specifically, the non-data-driven methods mostly originate from the period earlier than the era of machine learning. This is explained by considering that the initial studies were conducted with limited ultrasound data and hardware support. Later on, driven by the rapid development of computing hardware and software, such as faster GPUs with larger memories, advanced US imaging systems and open-source machine learning toolboxes, instrument detection solutions have been converging to data-driven methods, which are commonly denoted as machine learning.

A. Non-data-driven Methods

As for preliminary non-data-driven studies in instrument detection, there are two major classes based on a-priori knowledge of the instrument in US, which are called parametric space methods and physical space methods. More specifically, the parametric space method detects the instrument by applying some spatial transformation, i.e. spatial transformation from physical space to parameter space, on an intensity or thresholded image with prior knowledge of the tool. As for the physical space method, it performs mathematical or geometrical modeling in the physical space with the standard spatial coordinate system.

1) Parametric space methods: The methods of this category apply a spatial parametric transformation on US images with prior shape knowledge of the instruments, assuming e.g. a straight or curved line in 2D or 3D space. With the prior shape information, the instrument has strong response after some projection-based spatial parametric transformation, which accumulates the pixel or voxel intensities along the instrument propagation with respect to spatial location and direction. Here, it is assumed that the instruments yield a higher intensity value than background in the B-mode or thresholded US images. Ding *et al.* [15], [16] proposed to segment a needle in 3D US by applying spatial projection from 3D to 2D thresholded volumetric data, which iteratively adjusts the projection direction in 3D space to minimize the projected needle area in the 2D plane. Okazawa *et al.* [17] introduced the use of the 2D Hough Transformation with prior knowledge of the needle insertion angle. The pixel values along the estimated direction are accumulated to generate a histogram of the projected voxels, which produces the corrected needle direction with post-processing, including iteratively direction rotation, points rejections, etc. Zhou *et al.* [18], [19] and Qiu *et al.* [20] presented a 3D Hough Transformation-based method on a thresholded volumetric image, which selects the highest accumulated values in the transformed space as the spatial parameters of the needle. Similarly, Radon Transformation-based Parallel Integral Projection [21] on the voxel intensity was introduced by Barva *et al.* to detect the straight electrode in 3D US images. The instrument is detected as the maximized response point in the Radon parametric space, which accumulates the voxel intensity values along the propagation of the instrument, as shown in Fig. 2. However, the essential insight of their methods is similar to the Hough Transformation-based approaches, except for the case of thresholding. Later on, the Hough Transformation method was also applied on the 2D images, by projecting 3D images using a ray-casting approach [22]. Recently, Beigi and Rohling have employed temporal in-

formation to enhance the ability of the Hough Transformation [23], which detects the needle in the 2D+T(time) US images. Moreover, Daoud *et al.* have proposed to apply the 2D Hough Transformation technique to needle localization in 2D B-mode [24] images, where they also introduced Power Doppler as extra information, improving the performance over only B-mode methods.

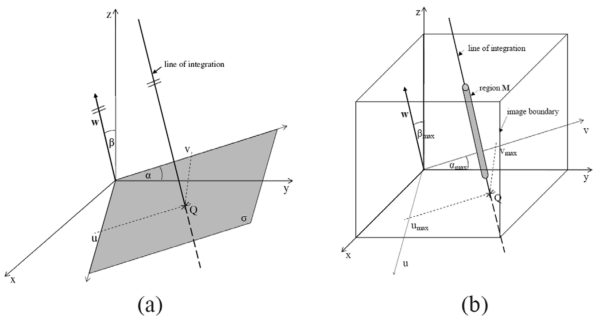

Fig. 2. Overview of Parallel Integral Projection (PIP) method by Barva *et al.* [21]. (a) PIP transform: the integral of the image intensity is calculated along a line given by the point $Q = [u, v]$ and a vector $\mathbf{w}$. (b) Evaluation of the PIP transformation of a 3-D image. The electrode is represented by a cylinder M. In this particular configuration, the parallel integral is maximized when the line of integration is identical with the electrode axis.

Above methods have some prior knowledges for the instrument detection: (1) instruments are straight or little bit curved in the images, which can be modeled by accumulating the values along the instrument axis through spatial transformation; (2) the instrument has higher intensity values than the background such that a simple threshold or intensity-based transformation can be directly applied to detect the target. With the previous assumptions, most methods were validated based on computer simulations [21] or phantom environment (*in-vitro*) [15], [17], [18], [19], [20], but in most works the challenges and difficulties were underestimated for instrument detection in noisy B-mode images. Although there were some studies [15], [18], [20], [23] that considered more challenging datasets from isolated tissue or even patient data, i.e. *ex-vivo* and *in-vivo* datasets, the simplified voxel thresholding without sufficient local or contextual information hampers the capacity of the detection algorithms by introducing too much outliers or giving an under-segmented instrument. In contrast, instrument model-fitting methods with proper pre-processing could be more suitable to model the instrument in B-mode images and improve the detection performance [25].

2) Physical space methods: The preliminary study for instrument modeling can be traced back to Novotny *et al.* [26], since they modeled the catheter as a voxel cluster with longest and straightest connectivity groups, which was implemented based on Principle Component Analysis (PCA). With this initial study, a texture-based instrument segmentation method [27] was proposed in 3D US images by applying Expectation-maximization, local texture analysis and PCA, which iteratively segments the instrument from an *in-vitro* dataset. Zhao *et al.* introduced 3D gradient orientation to calculate the instrument phase information [28], which segments the needle by applying Line-Support-Region analysis for grouped regions. Similar to this local gradient analysis, a histogram analysis method was designed by McSweeney *et al.* [29] to threshold the US image, which localizes the needle by morphological operation and line fitting. The Frangi vesselness filter [30] was considered to better describe the instrument and filter out instrument-related points in US images, based on assumptions on the high contrast of instrument edges compared to background and a tubular structure [31], [32], [33], [34]. Although these methods may include different pre-/post-processing steps, the core idea of them is to extract tubular-like structures by Hessian matrix analysis for local intensity distributions. Further processing steps range from a simple thresholding [31], [32], [35] to Random sample consensus (RANSAC) model-fitting [34] with Kalman filtering in time sequence-based US datasets [36], [33]. Moreover, the local Hessian matrix is also applied to detect the shadow of steep needles in 3D US [37], which automatically extracts the 2D slice containing the needle for in-plane visualization.

Beside the above vesselness filter-based methods, template matching with a pre-defined catheter filter was proposed by Cao *et al.* [38], which considered a 3D catheter template for candidate voxel selection. The resulting images were optimized by a likelihood map with shape measurement. Similarly, automatically optimized Gabor filter methods [39], [40], [41], [42], [43] were used with different image processing steps for needle segmentation in 2D US. Specifically, Kaya *et al.* [39], [40], [41] proposed to employ a two-stage method for needle localization based on Gabor filtering with an optimized insertion angle estimation. First, the Otsu's method is applied to obtain the binary image. Then, the needle in the binarized image is localized by RANSAC model-fitting, which generates the region-of-interest (ROI) for needle-tip probability mapping and localizes the tip. Their methods were validated with static images [39] and real-time video [40]. Moreover, they further implemented a simulation platform for needle tracking [41] for real-time localization. In contrast to Kaya *et al.* methods with complex post-processing, Hacihaliloglu *et al.* [42] employed log-Gabor filters to extract phase-symmetry information, which automatically selects the scale, bandwidth and orientation parameters to enhance the contrast of the needle. The needle is finally detected by a modified Maximum Likelihood Estimation SAmple Consensus (MLESAC) method [44]. Furthermore, Mwikirize *et al.* [43] proposed to localize the needle by introducing signal transmission maps for 2D US, which firstly enhances the visibility of the needle in noisy US images. Then, the needle is localized by applying the algorithm from Hacihaliloglu *et al.* [42].

In contrast to the above methods for static US images, there are some papers focusing on exploiting temporal information. Kaya *et al.* [46] proposed tracking the needle tip by applying a dynamic updated template to 2D US video, which measures the similarity between the template and US images to identify the target. This method avoids needle localization algorithms in each video frame, but requires template definition. Beigi *et al.* [47], [45], [48] intensively studied the needle detection by applying spectral analysis, which makes use of spatiotemporal information from natural hand tremor. This periodic pattern is hardly observed by human eyes, but can be captured by

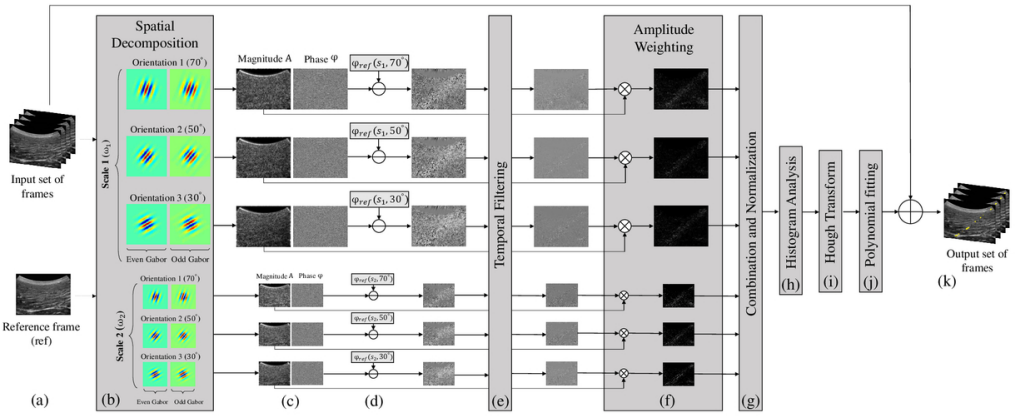

Fig. 3. Block diagram of Beigi *et al.* [45]. A reference frame is selected from the input sequence of the B-mode data and all frames are sent as the input to the algorithm (a). Three complex Gabor wavelet pairs form the steerable pyramid for spatial decomposition of the sequence (b) into local magnitude and phase measurements (c). Phase differences of all frames will be computed from the reference frame and (e) temporally filtered using a bandpass filter. Amplitude weighting is then performed on the filtered phase differences for adjustments in cases where magnitude responses are weak (f). Results from all scales and orientations are combined (g) and thresholded to generate the binary mask for the Hough transform (h). The Hough transform derives an estimate of the trajectory and discards some of the outliers (i). Model fitting is finally used to remove any remaining outliers and improve the trajectory detection (j). The detected needle is then added to the input sequence as an overlay (k).

spectral analysis of 2D B-mode images, which leads to a better result than static images. However, these image modalities were limited to 2D+T format, due to hardware constraints and complex filtering steps or real-time requirements, as is depicted in Fig. 3. Although there is a recent study focusing on 3D volumetric data with temporal information [49], they considered an extra camera for giving support information, which obtained a better performance than ROI-based Kalman filtering.

The above methods mainly follow an algorithm pipeline, which we denote as *segmentation-modeling* pipeline. First, carefully designed filters or instrument templates are applied to extract/enhance the instrument-related information in the image. Second, optimized thresholding is applied to binarize the images to segment the instrument from the data. Third, model-fitting algorithms in 2D/3D images are applied to localize the target. Although the processing steps can be different, most of above methods indicated that a successful segmentation method is the key step to detect the instrument in challenging US images, which significantly relates to the first and second steps. However, these segmentation methods are limited by prior knowledge of the instrument and is sensitive to image modality or appearance. Moreover, a simple thresholding with prior or empirical knowledge also limits the segmentation performances in different application cases. To better describe the instrument-related information and obtain more accurate segmentation results, data-driven methods were exploited as they can better describe the information with knowledge learned from the data itself.

B. Data-driven Methods

In recent years, with the fast development of hardware for data recording and processing, data-driven methods, also called machine learning methods, have been intensively exploited in computer vision and medical imaging analysis areas. The main idea of data-driven methods is to model task-related information by designing a proper mathematical representation from the training dataset, e.g. feature vector and pre-trained classifier, which is then used to make a prediction or decision without being explicitly programmed [50]. There are two popular and widely studied approaches, the first one is handcrafted feature design with a machine learning classifier. This method employs feature vector extraction and task classification. However, the design of handcrafted features requires task-related knowledge and experience, which hampers the classification performance and therefore is gradually replaced by a recently developed learning technology, i.e. deep learning. Deep learning is a fully data-driven method, which combines feature extraction and classification with a fully automated information learning style. Deep learning methods can automatically learn the task-related information from provided data and, in most cases, learn more powerful feature representations than the previous handcrafted feature design methods.

1) Handcrafted feature-based methods: Krefting *et al.* [52] proposed a multi-threshold-based needle detection method in video sequence by applying contextual statistical features. For each frame of the sequence, multi-thresholding is applied to obtain the binarized images for feature extraction, which are then classified by a Mahalanobis distance-based linear classifier. In contrast, a common approach to detect the instrument by machine learning methods is to extract a feature vector for each pixel or voxel in the US image, which is then classified by a pre-trained supervised machine learning model. The model assigns the category for each point, which generates the segmented image for the post-processing, e.g. RANSAC model-fitting or Radon/Hough transformation.

Because machine learning methods provide better segmentation performance, they have been intensively studied in recent years. Uherčík *et al.* [53] proposed to use voxel intensity, Frangi vesselness response and axis descriptors as the discriminating features for needle voxel classification by a Cascade classifier, which obtained a much higher successful rate and

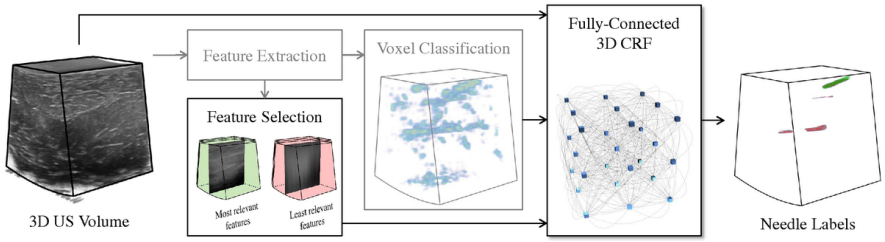

Fig. 4. Block diagram of CRF-based needle detection [51]. The feature vectors for each voxel are extracted and selected for voxel classification and fully-connected 3D CRF, respectively. With the aid of initial voxel classification, the 3D CRF processes contextual correlations between selected features and the segmented volume.

accuracy than Parallel Integral Projection and Random Hough Transformation methods. Hatt *et al.* [54] proposed to consider second-order Gaussian derivative filters for producing the features describing pixels in 2D US images, which are classified by AdaBoost to segment needle. The Radom Transformation is applied to localize the instrument in the image. Their results demonstrated a better performance than simple and straightforward thresholding methods like intensity thresholding, filtered thresholding or Frangi vesselness thresholding.

Pourtaherian *et al.* extensively studied needle detection by applying 3D orientation-invariant Gabor features with Linear Support Vector Machine (LSVM) [55], which is processed by a RANSAC algorithm to localize the instrument [56], [57], [58]. Their studies show that the Gabor filter with LSVM could properly capture the spatial information for a long but thin instrument in complex 3D US images. However, their initial studies for catheter segmentation were only validated in an *in-vitro* dataset with insufficient segmentation results for accurate detection. Later on, Zanjani *et al.* [51] demonstrated the Gabor features can be simplified by feature selection and the segmentation performances can be boosted by Conditional Random Field (CRF) [59] than a simple LSVM, which segments the images based on contextual-level information correlation (as shown in Fig. 4). To further exploit complex anatomical information in 3D US for catheter segmentation in cardiac catheterization, Yang *et al.* proposed multi-scale and multi-definition features for supervised learning classifiers, which demonstrated a better discriminating information extraction than techniques solely based on Gabor features [60], [61]. The segmented instrument was fitted by a more complex Sparse-Plus-Dense RANSAC algorithm to fit the curvature instrument in the cardiac chambers. Similar to Pourtaherian, Mwikirize *et al.* [62] also considered log-Gabor features as the local phase extractor, which is processed by locally normalized histogram of orientated gradients (HOG) features, to describe the needle in 2D US slices. The constructed HOGs are then classified by LSVM to segment and enhance the needle in 2D US slices. In the mean time, Younes *et al.* [63] proposed to make use of Gaussian mixture model-based Naive Bayes classifier to segment the needle in 3D prostate brachytherapy US. The needle in US is finally localized by a standard RANSAC model-fitting.

The above methods are applied on static images, as the voxel-level feature extraction and classification in US images are computationally expensive for current hardware. Nevertheless, still several papers focusing on 2D US with temporal information can be found. Beigi *et al.* [64] proposed to detect a hardly visible needle in 2D+T US by applying local phase extraction with temporal sequence analysis. More specifically, the phase information for each frame is extracted to formulate the element from a time-sequence-based phase video, which is then processed by Auto-Regressive Moving-Average (ARMA) model to extract the feature vector. With classification from a modified SVM, the small motion of the invisible needle can be characterized for needle detection. Furthermore, Beigi *et al.* employed spectral feature analysis on spatiotemporal features derived from optical-flow analysis [65], which allows to detect and track the needle in a 2D US video. In contrast to these off-line learning methods, Mathiassen *et al.* [66] applied on-line learning, i.e. learning and updating the needle-related information during the video progressing. They applied the statistical filtering methods, i.e. Kalman filter and Particle filter, to learn the appearance of the instrument in the video with real-time performance.

Beside the above machine learning with classification, Zhang *et al.* [67], [68] proposed to detect multi-needle in 3D brachytherapy by unsupervised sparse dictionary learning (SDL). Specifically, the needles and tissue information in the 3D images are encoded into latent space, which are then distinguished by the sparse dictionary model. Based on the SDL, the needle in the 3D space can be captured and reconstructed in the volume that multiple needles can be localized by region-of-interest-based RANSAC model-fitting.

Even though the above methods achieved satisfactory detection results for the given tasks, it is difficult to design the optimal feature representation, which is the key reason that hampers the segmentation. This leads to complex post-processing to avoid the outliers or false positives. Moreover, the designed features can only focus on local information, while ignoring the contextual and semantic information [51]. To handle these limitations, deep learning methods based on neural networks were proposed and studied in recent years [69].

2) Deep learning based methods: At the beginning of the era of deep learning, Geraldes and Rocha proposed to consider neural networks, i.e. Multilayer Perceptron network (MLP), to segment the needle in 2D US images [71], [72]. The segmented results are used to guide the Kalman filter to

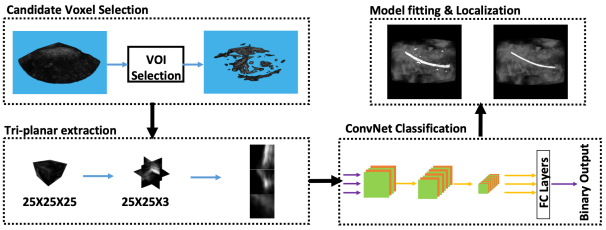

Fig. 5. Block diagram of VOI-based CNN for catheter detection [70]. The input volume is first processed by a Frangi filter to select the VOI voxels, which are then classified by a tri-planar-based CNN for voxel-based classification. The RANSAC model-fitting is applied to localize the catheter in 3D B-mode images.

track the needle tip in the video sequences [71]. These papers demonstrated the feasibility of deep learning to detect the a medical instrument in challenging US images.

The conventional voxel-based classification in previous studies was extended into deep learning methods, which employed convolutional neural networks (CNN) to classify the voxels' category by a classification strategy on whole 3D images. Pourtaherian *et al.* [73], [74] and Yang *et al.* [75] proposed a tri-planar CNN method to segment the needle or catheter in 3D US volume, which simplified the computational cost by extracting three orthogonal slices of 3D local patches instead of a full 3D patch. Nevertheless, this exhaustive strategy is time-consuming because it applies a CNN on every point of the image iteratively, which cannot satisfy real-time clinical usage. Later on, this exhaustive strategy was overcome by applying the Frangi vesselness filter as a voxel-of-interest (VOI) pre-selection step [70], which is shown in Fig. 5.

The voxel-based classification is time-consuming and can be easily limited by large-sized US images, which are processed in the sample domain [76]. As a consequence, a fast region-based CNN (Fast R-CNN) is combined with regional proposal network (RPN) to efficiently detect the needle in 2D US images [77]. Specifically, the Fast R-CNN is considered to generate the shared feature maps for RPN, which classifies and regresses the location of the needle in the input image. By doing so, the location of the needle in 2D US images is annotated by a bounding box. However, this method cannot accurately segment and localize the instrument skeleton at pixel or voxel level. To overcome this limited performance and leverage the powerful fully convolutional network (FCN), which assigns the class categories to all the points of the input image by using semantic information, semantic segmentation was introduced and studied for instrument segmentation.

To semantically segment the instrument in US, an FCN with U-Net structure [78] was considered as a solution, since it exploits the semantic information at different image scales with skipping connections for data flow. This approach leads to state-of-the-art performance in most applications in the medical imaging area (means limited training images for the network). A 2D FCN is applied to segment the needle in 2D US [79], [80] and in 3D US by applying spatial decomposition, i.e. decomposing the volume into stacks of slices [81], [82]. However, in terms of 3D volumetric data, the decomposition approach limits the semantic information usage due to the compromised 3D information after slicing. To address this limitation, patch-based 2.5D or 3D UNet were proposed to segment the cardiac catheter [83], [84], [85] or prostate needles [86] in 3D volumetric data by dividing the image into smaller patches, which preserves the 3D contextual information and clearly reduces the GPU memory requirements for 3D deep learning. Nevertheless, this patch-based strategy limits the whole image contextual information usage. To overcome this limitation, Arif *et al.* applied an extremely simplified 3D UNet on a complete 3D image to segment the needle in ultrasound imaging for liver [87]. Although their method showed a successful result in that application, its generalization and segmentation abilities are constrained by the simplified network design [85].

Due to requirements of a large amount of datasets and high GPU memory usage for deep learning, temporal information is not widely investigated in deep learning-based medical instrument detection. Mwikirize *et al.* [88], [89] proposed time-difference-based regression and classification CNNs to detect the needle in 2D US sequences. The differences between two adjacent frames are obtained by applying pixel-wise logical operation, which captures subtle motion of the needle and feeds it into CNNs for the detection. Nevertheless, these methods process the temporal information outside the CNN such that the spatial-temporal information may not be properly handled by the deep learning approach.

In contrast to commonly used CNNs, convolutional dictionary learning was proposed by Zhang *et al.* [90] as an extension from their work [67]. Instead of their previous work using sparse dictionary learning (SDL), Zhang *et al.* considered a convolutional sparse coding model to replace the SDL method [90], which used CT images as the a supervisory signal to create the dictionary for reconstructing the detected needles in the 3D volumetric data.

Although the deep learning methods provide advantages in segmentation accuracy and better information description, this data-driven method requires a large amount of training data with annotations that hampers the real clinical usage. Moreover, most state-of-the-art methods are far from real-time performance, so that optimizations are still required. These limitations form also key issues for employing deep learning when considering mature clinical applications.

III. Clinical Applications

This section discusses the main clinical applications related to medical instrument detection. More specifically, up to now, the paper has mainly categorized current papers into four classes with clearly defined clinical applications being validated on clinical datasets, which are: (1) US-guided anesthesia-based intervention, (2) US-guided biopsy therapy, (3) US-guided prostate brachytherapy and (4) US-guided cardiac catheterization.

A. US-guided anesthesia-based intervention

Needle-based regional anesthesia or imposing blockade is important in current clinical practice, which provides a safer

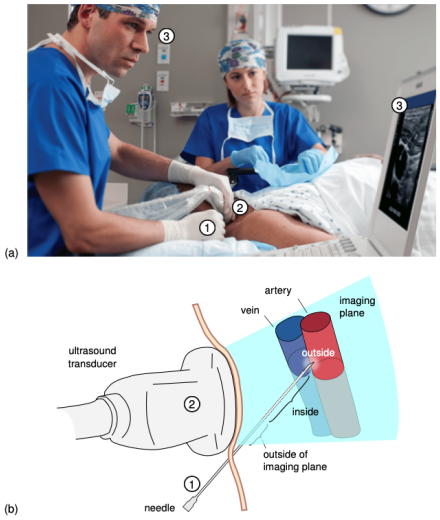

Fig. 6. Utilization of US to guide needle therapy [91] (a) The clinical situation showing the multi-fold coordination of ① the needle, ② ultrasound transducer, while ③ looking at the US screen (Courtesy of Philips Ultrasound). (b) Schematic representation of guiding a needle using US imaging, depicting an example situation for regional anesthesia, where the needle tip is outside the imaging plane and is approaching an erroneous target area, as a result of poor needle-transducer alignment.

and more accurate intervention for further procedures. Conventional regional anesthesia requires experienced surgeons to deliver the medicine to the correct region, which is commonly guided by ultrasound imaging, since it provides a fast and convenient visualization solution for clinical experts. However, as shown in Fig. 6, multi-fold coordination of the US screen, needle and ultrasound probe complicates the procedure and hampers the operation outcomes with higher risks. As a result, extensive training for a surgeon is required to achieve a successful therapy under the guidance of the ultrasound imaging.

To visualize the needle during the US-guided regional anesthesia or blockade, an essential constraint should be satisfied in conventional 2D US: the needle should be positioned in-plane in 2D images, where the needle is visualized as a bright line, requiring a perfect alignment between instrument and ultrasound plane [43]. However, this 2D US-guided therapy is facing the challenges of the instrument being invisible [64], or the instrument being out of the plane [58]. As a consequence, 3D volumetric ultrasound is gradually adopted into clinical usage, because it can provide richer spatial information of the needle. However, complex 3D information and complicated 3D image visualization hampers the efficiency of the surgeons when they are looking for the needle and guiding it to the target region. As a consequence, automatic needle detection is investigated to facilitate the clinical interventions and improve the operation outcomes. With extensive validations in [58], 3D US-guided needle anesthesia have been proven to be a promising solution to facilitate this type of regional intervention.

B. US-guided biopsy

Biopsy of nodule or lymph node is essential for diagnosis, particularly for finding malignant tissue. To obtain the tissue

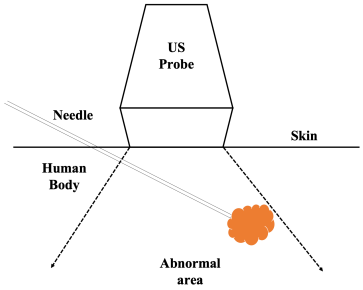

Fig. 7. Example of US-guided needle biopsy. With guidance of US imaging, the needle is correctly placed to abnormal regions for performing a biopsy, such as breast biopsy or thyroid nodule biopsy.

samples by biopsy, needle biopsy or open surgical biopsy is commonly considered, based on suspected pathology, patient health condition and procedure complexity. Although conventional open surgery biopsy provides better diagnostic results, the less invasive needle-based biopsy becomes attractive, as it offers a reasonable result. Historically, needle-based biopsy was performed by radiologists in special procedure rooms with interventional radiology suites, which is however gradually replaced by a US imaging system because of its lower cost, higher healthcare efficiency and better tissue characterization. However, the drawbacks of US imaging need to be addressed, like difficult interpretation, lower image contrast than traditional X-ray imaging, and the fact that extra training is required for surgeons to obtain a satisfied tissue sample [92].

Similar to the above regional anesthesia, multi-dimensional coordination has complicated the procedure of the needle guidance. As a result, automatic instrument detection is necessary to help surgeons in performing their tasks. Moreover, US imaging is also increasingly used for tissue characterization or abnormality detection, which can facilitate the biopsy procedures together with instrument detection. With a richer 3D spatial information, 3D US can efficiently and accurately facilitate the surgeons to perform the operations and reduce the risk for patients. As has been proven by Arif *et al.* [87], 3D US-guided liver biopsy can facilitate needle detection in volumetric data.

C. US-guided prostate brachytherapy

Prostate cancer is the development cancer in the prostate, which is the second-most common cancer in male patients. Prostate therapies are important treatment worldwide for prostate cancer. Specifically, prostate brachytherapy is a highly recommended treatment for patients at the early stage of cancer development. To perform the therapy, the needles or catheters are used to place radioactive particles (so-called seeds), which have the size of a grain of rice, to the tumor regions. It delivers a high-dose radiation (HDR) to the tumor without affecting the normal tissue around abnormal areas. To guide the instrument to perform seeds implantation in the prostate, Transrectal Ultrasound (TRUS) has been considered since the 1980s [93], which facilitates the operation and results in better experience for patients.

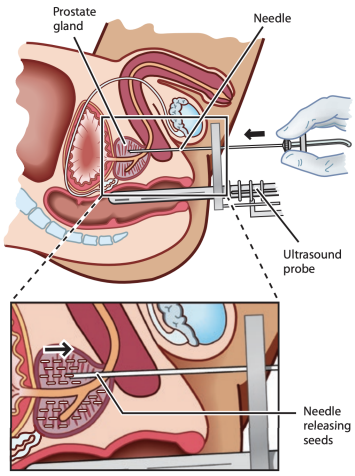

Fig. 8. Example of US-guided prostate brachytherapy. Needles are guided by US images to place radiation seeds for prostate cancer (source: Understanding Brachytherapy for Prostate Cancer from https://www.prostate.org.au/).

In contrast to other applications, prostate brachytherapy requests to insert multiple instruments into the prostate, so that multi-instrument detection in TRUS is essential for a successful operation planning. However, because of this requirement and instrument placement condition, i.e. needles are close to each other, the detection algorithms need to be stable and accurate enough for the detection of multiple objects, which is rarely studied in state-of-the-art solutions. Moreover, beside this multi-detection challenge, an efficient detection algorithm is required because the common operation time is around 90 minutes for prostate brachytherapy [93]. As shown in Zhang *et al.* [68], the needle detection procedures are extremely accelerated to a half minute, whereas conventional needle digitization takes around 15-20 minutes by an experienced physician.

D. US-guided cardiac catheterization

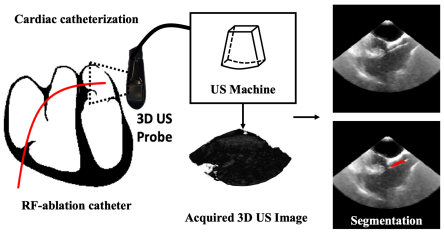

Fig. 9. Example of cardiac catheterization under US guidance. The ultrasound probe captures the 3D volumetric data of the heart chamber, which contains an RF-ablation catheter. The algorithm outputs the segmentation results and automatically demonstrates the results in 2D slices for a better view.

Similar to US-guided needle therapy, US-guided catheterizations, especially for cardiac applications, are also facing the challenge of multi-dimensional coordination for surgeons. Moreover, due to complex anatomical structures in the heart chambers, sonographers need more time to localize the less obvious catheter (compared to the metal needle) in 3D volumetric data by a slice-by-slice tuning procedure, which is time-consuming and complicates the operation. Nevertheless, 3D ultrasound is attractive for catheterizations because of its radiation-free nature and easy-to-use properties, and its richer offering of spatial information for tissues [61]. As a result, it is a promising choice to support or replace current X-ray imaging for cardiac interventions. An example of cardiac catheter detection in 3D volumetric data is shown in Fig. 9.

To detect the catheter in 3D US images, several solutions were proposed [38], [61], [85], which employed different methods as reviewed in Section II. However, there are still several challenges that need to be addressed in the future work: (1) US-guided catheterization is not widely accepted for clinical usage so that expert knowledge is limited by experience. Furthermore, a clear definition is lacking of how accurate and how fast the instrument detection should be. (2) Due to hardware limitations of the US probe, the 3D volumetric data only focuses on small field-of-view of the heart chamber, which then lacks of the guidance capability when the instrument is inserted into the target region, so that a cardiac catheter still requires X-ray to perform guidance in the vein. This hampers the flexibility of US-guided interventions. (3) Only sets of limited studies has been performed on US-guided catheterizations when compared to needle-based interventions, which makes US-guided catheterizations a minority area in the field of computer-assisted interventions.

IV. Evaluation Dataset, Metrics and Results

In this section, the papers are summarized based on the above detection methods. To complete the discussions on these papers, we summarize the information w.r.t. datasets, like data type, data format and target clinical application, which is covered in overview by Table I and Table II. More specifically, as for dataset type, there are four different categories: simulation dataset, *in-vitro* dataset, *ex-vivo* dataset and *in-vivo* dataset. These types of dataset are explicitly defined and explained as follows.

Simulation: Computer software-based simulation was used to generate ultrasound images with well-defined instrument information with image content. Filed II [94] is the commonly used simulation platform for US image generation. Because it can provide sufficient image results without extra cost for US equipment and experiment labs, such as biological lab for tissue experiments and target instrument operations, which require extra clinical doctors to guide the experiments. As a result, a simulation dataset can be used to validate the scientific idea and perform a feasibility study. However, this approach lacks complex anatomical structures of the tissue, which limits the clinical value and stability of the methods.

In-vitro: In contrast to simulation datasets, *in-vitro* datasets provide a more realistic case, which employ polyvinyl alcohol (PVA) or agar phantoms to mimic the human tissue. Moreover, these datasets were recorded from real US imaging equipment with proper post-processing, which tried to mimic real clinical scenarios. However, this approach has an essential limitation that the phantom cannot include complex and detailed anatomical structures in the real clinical applications, such as vessels or muscles. Moreover, due to different physical properties

between a phantom and real tissue, the US imaging results are also different, which limits the robustness and clinical value of the algorithms.

Ex-vivo: Similar to *in-vitro* data, *ex-vivo* data is another commonly used dataset type, which replaces PVA or agar phantoms to an isolated real animal tissue, such as chicken breast for needle detection or porcine heart for catheter detection. These kinds of tissues provide a more complex image appearance due to anatomical structures, which are much more similar to clinical applications. Nevertheless, this dataset type still has limitations of the recording conditions, such as less complex muscle and vessel structure of chicken breast when considering a needle detection for anesthesia or water-filled (rather than real blood) heart chamber for cardiac catheterization. Although these limitations hamper the clinical value, they provide a more stable and promising comparison when compared to the above non-tissue-based datasets. Because of the data recording difficulties, such as biological experiment certification and support from clinical experts, *ex-vivo* datasets are still important and are considered in a majority of recent papers, which indicates its importance for algorithm validation.

In-vivo: In this article, *in-vivo* data is defined as the dataset coming from real clinical operation, such as prostate biopsy, or live animal dataset, e.g. live porcine for cardiac catheterization. Because the datasets were recorded from real clinical usage or mimicking it, these *in-vivo* datasets demonstrated important clinical value for the algorithm validation. However, recording of this type of dataset is challenging when compared to the above three dataset types. First, it is difficult to ask surgeons to record the data when the therapy is not widely accepted in the hospital, or data recording can hamper the operation procedures. Second, due to privacy policy, it is difficult to obtain the patient dataset without complex anonymization steps. Third, the cooperation with clinical experts can be difficult, because various surgeons have different preferences for image appearance and configurations, which makes it complicated to agree on a unified image quality and appearance for algorithm development.

As summarized in Table I and Table II, only the most valuable clinical datasets are reported, i.e. *in-vivo* or *ex-vivo* datasets, and are indicated for multi-dataset validation papers.

With the above definitions of the datasets, algorithms were validated and summarized in Table III and Table IV. To evaluate the detection performances, several validation methods are used and summarized: (1) non-learning-based methods are commonly validated on the dataset as described, i.e. validate the method on the whole dataset; (2) learning-based methods are commonly validated on a testing dataset after using a different but similar training dataset for model training. More specifically, cross-validation like leave-one-out cross-validation (LOOCV), k-fold cross-validation (k-CV) or a straightforward dataset split (S), i.e. dividing the dataset into training, validation and testing subsets, are commonly used in learning-based methods.

As for evaluation metrics, the orientation error (OE, instrument-axis direction mismatch), tip error (TE, instrument-tip mismatch), detection error (DE, location-detection error), end-points error (EE, average mismatch of instrument tip and tail), diameter mismatch (DM), tip-to-plane error (PE, point-plane distances between the end-points of the ground-truth needle and the detected plane) and axis error (AE, average error of each point on the instrument axis) are used for localization or detection accuracy. Precision, Recall and Dice score (DSC) are used to measure the segmentation performance at pixel or voxel level. Time efficiency is denoted as t for simplicity. Hough/Radon Transformations are simplified as Hough/Radon. KF and PF represents Kalman filter and Particle filter, respectively. Other notations or abbreviations are defined in the above sections. It is worth to mention that the spatial resolution of the images are not summarized because some of the papers did not include this parameter. Some table cells are empty due to unclear descriptions from the considered papers.

Although there is no commonly used benchmarking dataset for a fair comparison, the detection performance from the literature shows a promising accuracy for clinical usage as they consider almost the same evaluation metrics. However, in terms of real-time efficiency, which is commonly required in clinical applications for operation guidance, it is far from real time in most papers. Specifically for papers validated on the *in-vivo* datasets, most of them are far from real clinical usage. Since most of them were validated on limited clinical data, e.g. around 10 patients or even less, without reporting a sufficient time efficiency (especially for 3D imaging). Nevertheless, all the detection results show promising accuracy for guidance. This result demonstrates a trend that the researchers in the area are more focusing on detection accuracy instead of real-time performance, especially for papers w.r.t. 3D volumetric datasets.

V. Discussion

To conclude this review, we have selected several representative papers for medical instrument detection in US images from literature. These papers are not claimed to be the 'best' or the most cited papers, but have been selected based on their detection methods for different approaches and/or whether evaluation was performed on clinical datasets. The order of the selected list below is ranked by year of publication.

Novotny *et al.* (2003) [26]: As one of the preliminary works for instrument detection in 3D US images, this paper considered a PCA-based shape prior knowledge to detect the instrument in thresholded images. It already included essential steps of the detection methods, which are followed and considered by later studies. Although the direct thresholding is simple and straightforward, the segmentation step shows the importance for the later post processing, such as PCA, Hough transformation or model-fitting algorithm. The PCA-based shape analysis indicates the importance to model the instrument shape information, i.e. a curved tube or line in the 3D space. The combination of two main steps are used and extended by later studies in different forms.

Zhou *et al.* (2007) [18]: In contrast to PCA-based model analysis, the Hough transformation can better capture the geometric information of the instrument, since it models the tool as the thicker line in the 3D space without a carefully tuned Eigenvalue analysis. The Hough transformation can better

TABLE I
OVERVIEW OF PAPERS USING NON-LEARNING-BASED TECHNIQUES FOR INSTRUMENT DETECTION, ORDERED BY YEAR OF PUBLICATION. CLINICAL APPLICATIONS ARE MENTIONED IN THE LITERATURE.

Reference	Year	Format	Application	Dataset
Ding *et al*. [15]	2002	3D	Needle biopsy/brachytherapy	*in-vitro/ex-vivo*
Novotny *et al*. [26]	2003	3D	Graspers for cardiac/fetal surgery	*ex-vivo*
Ding *et al*. [16]	2004	3D	Needle biopsy/brachytherapy	*in-vitro/ex-vivo*
Okazawa *et al*. [17]	2006	2D	Needle biopsy/drug delivery	*in-vitro*
Linguraru *et al*. [27]	2006	3D	Tntracardiac operation	*in-vitro/in-vivo*
Zhou *et al*. [18]	2007	3D	RF-ablation for uterine operation	*in-vitro*
Zhou *et al*. [19]	2008	3D	RF-ablation for uterine operation	*in-vitro*
Barva *et al*. [21]	2008	3D	Biopsy/neurological	simulation
Aboofazeli *et al*. [22]	2009	3D	Biopsy	*in-vitro*
Zhao *et al*. [28]	2009	3D	Biopsy	*in-vitro*
Ren *et al*. [31]	2011	3D	Catheter intervention	*in-vitro*
Zhao *et al*. [36]	2013	3D	Needle biopsy	simulation
Cao *et al*. [38]	2013	3D	Catheter for cardiac/prostate/biopsy	*in-vivo*
Mohareri *et al*. [32]	2013	3D+t	Prostate therapy	*in-vitro/ex-vivo/in-vivo*
Zhao *et al*. [33]	2013	3D+t	Needle biopsy	simulation
Qiu *et al*. [20]	2013	3D	prostate therapy	*in-vitro/ex-vivo/in-vivo*
Malekian *et al*. [34]	2014	3D	Catheter biopsy	simulation
McSweeney *et al*. [29]	2014	2D	Needle biopsy/nerve block	*in-vitro*
Kaya *et al*. [39]	2014	2D	Needle biopsy/drug delivery	*in-vitro*
Kaya *et al*. [40]	2014	2D	Needle biopsy/drug delivery	*in-vitro*
Beigi *et al*. [23]	2014	2D+t	Needle biopsy/anesthesia/delivery	*in-vitro/ex-vivo*
Beigi *et al*. [47]	2015	2D+t	Needle biopsy/anesthesia/delivery	*in-vitro/in-vivo*
Hacihaliloglu *et al*. [42]	2015	2D	Needle biopsy/anesthesia/therapy	*ex-vivo*
Kaya *et al*. [41]	2015	2D+t	Needle biopsy/drug delivery	*in-vitro*
Pourtaherian *et al*. [37]	2016	3D	Needle anesthesia/ablation	*ex-vivo*
Beigi *et al*. [45]	2016	2D+t	Needle biopsy/nerve block/anesthesias	*in-vitro/in-vivo*
Mwikirize *et al*. [43]	2016	2D	Needle biopsy/ablation/anesthesia	*ex-vivo*
Beigi *et al*. [48]	2016	2D+t	Needle biopsy/nerve block/anesthesia	*in-vivo*
Kaya *et al*. [46]	2016	2D+t	Needle biopsy/drug delivery	*in-vitro*
Daoud *et al*. [49]	2018	3D	Needle intervention	*ex-vivo*
Daoud *et al*. [24]	2018	2D	Needle intervention	*ex-vivo*
Agarwal *et al*. [35]	2019	2D+t	Anesthesia/biopsy/brachytherapy	*in-vitro*

TABLE II
OVERVIEW OF PAPERS USING LEARNING-BASED TECHNIQUES FOR INSTRUMENT DETECTION, ORDERED BY YEAR OF PUBLICATION. CLINICAL APPLICATIONS ARE MENTIONED IN THE LITERATURE.

Reference	Year	Format	Application	Dataset
Krefting *et al*. [52]	2007	2D	Prostate biopsy	*in-vivo*
Uherčík *et al*. [53]	2013	3D	Needle biopsy/brachytherapy	simulation/*in-vitro/in-vivo*
Geraldes *et al*. [71]	2014	2D+t	Needle anesthesia/biopsy/brachytherapy	*in-vitro*
Rocha *et al*. [72]	2014	2D	Needle anesthesia/biopsy/brachytherapy	*in-vitro*
Pourtaherian *et al*. [56]	2015	3D	Needle biopsy/ablation/anesthesia	*in-vitro/ex-vivo*
Hatt *et al*. [54]	2015	2D	Needle biopsy/nerve block/anesthesia	*ex-vivo/in-vivo*
Pourtaherian *et al*. [57]	2015	3D	Needle biopsy/anesthesia	*ex-vivo*
Mathiassen *et al*. [66]	2016	2D+t	Needle ablation/biopsy	*ex-vivo*
Pourtaherian *et al*. [58]	2017	3D	Needle/catheter intervention	*in-vitro/ex-vivo/in-vivo*
Beigi *et al*. [64]	2017	2D+t	Needle biopsy/ablation/anesthesia	*in-vivo*
Beigi *et al*. [65]	2017	2D+t	Needle biopsy/ablation/anesthesia	*in-vivo*
Mwikirize *et al*. [62]	2017	3D	Needle anesthesia	*ex-vivo*
Zanjani *et al*. [51]	2018	3D	Needle biopsy/anesthesia	*ex-vivo*
Mwikirize *et al*. [77]	2018	2D	Needle anesthesia/oncology	*ex-vivo*
Yang *et al*. [60]	2018	3D	Cardiac catheterization	*in-vitro/ex-vivo*
Younes *et al*. [63]	2018	3D	Prostate brachytherapy	*in-vivo*
Pourtaherian *et al*. [81]	2018	3D	Needle biopsy/ablation/anesthesia	*ex-vivo*
Pourtaherian *et al*. [74]	2018	3D	Needle biopsy/ablation/anesthesia	*ex-vivo*
Yang *et al*. [61]	2019	3D	Cardiac catheterization	*in-vitro/ex-vivo/in-vivo*
Yang *et al*. [70]	2019	3D	Cardiac catheterization	*ex-vivo*
Yang *et al*. [82]	2019	3D	Cardiac catheterization	*ex-vivo*
Yang *et al*. [83]	2019	3D	Cardiac catheterization	*ex-vivo*
Mwikirize *et al*. [89]	2019	2D+t	Needle biopsy/anesthesia	*in-vitro/ex-vivo*
Mwikirize *et al*. [88]	2019	2D+t	Needle biopsy/anesthesia	*ex-vivo*
Arif *et al*. [87]	2019	3D	Needle biopsy	*in-vitro/in-vivo*
Yang *et al*. [85]	2019	3D	Cardiac catheterization	*ex-vivo/in-vivo*
Min *et al*. [76]	2020	3D	Cardiac catheterization	*ex-vivo*
Rodgers *et al*. [80]	2020	2D/3D	Interstitial gynecologic brachytherapy	*in-vitro/in-vivo*
Zhang *et al*. [68], [67]	2020	3D	prostate brachytherapy	*in-vivo*
Zhang *et al*. [86]	2020	3D	Prostate brachytherapy	*in-vivo*
Zhang *et al*. [90]	2020	3D	Prostate brachytherapy	*in-vivo*
Lee *et al*. [79]	2020	2D	Needle biopsy	*in-vivo*

TABLE III
PERFORMANCES OF PAPERS USING NON-LEARNING-BASED TECHNIQUES FOR INSTRUMENT DETECTION. ORDERED BY YEAR TO MATCH TABLE I. - MEANS NOT REPORTED IN THE PAPER.

Reference	Year	Image Size	Size of Dataset	Key Metrics	Performance
Ding *et al*. [15]	2002	$357 \times 326 \times 352$	-	OE/t	1°/1-3 sec.
Novotny *et al*. [26]	2003	$128 \times 160 \times 64$	-	-	-
Ding *et al*. [16]	2004	$357 \times 326 \times 352$	6 volumes	EE/OE/t	0.7 mm/1.2°/13FPS
Okazawa *et al*. [17]	2006	482×398	10 images	AE	0.2-0.8 mm
Linguraru *et al*. [27]	2006	-	-	-	-
Zhou *et al*. [18]	2007	$381 \times 381 \times 250$	100 trails	OE/EE/t	1.93°/2.03 mm/0.22 sec.
Zhou *et al*. [19]	2008	$381 \times 381 \times 250$	100 trails	OE/EE/t	1.58°/1.76 mm/1.76 sec.
Barva *et al*. [21]	2008	$53 \times 71 \times 3100$	8 volumes	AE/TE/t	0.301 mm/0.263 mm/1121 sec.
Aboofazeli *et al*. [22]	2009	$256 \times 256 \times 125$	15 volumes	TE/t	2.8 mm/3 sec.
Zhao *et al*. [28]	2009	$50 \times 50 \times 50$	6	AE/DE/t	$< 2^\circ$/<2 voxel/$\sim$1.93 sec.
Ren *et al*. [31]	2011	-	6 trails	DM	< 0.4 mm
Zhao *et al*. [36]	2013	-	-	TE/AE/OE improvement	$> 92\%/ > 72\% > 71\%$
Cao *et al*. [38]	2013	$180 \times 130 \times 35$	26 volumes	TE/t	1.11 mm/0.41 sec.
Mohareri *et al*. [32]	2013	-	12	target registration error/t	2.68 mm/5 min.
Zhao *et al*. [33]	2013	-	-	AE/OE improvement	$> 60\% > 63\%$
Qiu *et al*. [20]	2013	$264 \times 376 \times 630$	40 volumes	OE/TE	0.8°/1 mm
Malekian *et al*. [34]	2014	$53 \times 71 \times 160$	28 volumes	failure percent/t	0-70%/<10 sec.
McSweeney *et al*. [29]	2014	-	-	-	-
Kaya *et al*. [39]	2014	640×480	723 images	detection rate/t	100%/0.31 sec.
Kaya *et al*. [40]	2014	640×480	164 images	detection rate/t	100%/0.234 sec.
Beigi *et al*. [23]	2014	-	-	TE	0.16-5.66 mm
Beigi *et al*. [47]	2015	2.5×2.5 cm^2	30 images	TE	0.5-0.7 mm
Hacihaliloglu *et al*. [42]	2015	450×450	150 images	TE/t	0.49-0.53 mm/0.8 sec.
Kaya *et al*. [41]	2015	640×480	112/54/38 images	OE/TE/t	1.95°/1.22/< 17 ms.
Pourtaherian *et al*. [37]	2016	-	100 trails	OE/PE/t	3.44°/0.66 mm/6.25 sec.
Beigi *et al*. [45]	2016	-	20 trails	OE/TE	0.93°/1.53 mm
Mwikirize *et al*. [43]	2016	370×370	100 images	DE/t	0.3 mm/0.6 sec.
Beigi *et al*. [48]	2016	-	20 sequences	OE	2.83°
Kaya *et al*. [46]	2016	640×480	7074 images	t	0.10-0.24 sec.
Daoud *et al*. [49]	2018	-	450/45 trails	OE/AE	$3.2 - 4.6^\circ$/4.0-4.4 mm
Daoud *et al*. [24]	2018	-	117 images	OE/AE/TE	$0.2 - 0.8^\circ$/0.2-0.6 mm/0.3-0.6mm
Agarwal *et al*. [35]	2019	-	$\sim$160 frames	TE	0.598 mm

TABLE IV
PERFORMANCES OF PAPERS USING LEARNING-BASED TECHNIQUES FOR INSTRUMENT DETECTION. ORDERED BY YEAR TO MATCH TABLE II. - MEANS NOT REPORTED IN THE PAPER. $\sim$ MEANS AVERAGE ESTIMATION OF IMAGES. N IS A VALUE BASED ON PATIENTS

Reference	Year	Image Size	Size of Dataset	Key Metrics	Performance
Krefting *et al*. [52]	2007	-	S:1650/-/4950 images	failure rate	6%
Uherčík *et al*. [53]	2013	$273 \times 383 \times 208$	S:18/-/3 volumes	failure rate/EE/t	0%/<0.5 mm/$\sim$ 300 sec.
Geraldes *et al*. [71]	2014	-	S:1335/-/422 images	DE	5.68-39.8 mm
Rocha *et al*. [72]	2014	101×101	S:2272/568/710 images	mean square error	0.0066
Pourtaherian *et al*. [56]	2015	$\sim 148 \times 169 \times 159$	LOOCV: 4/4 volumes	DE/OE/t	0.65-0.9 mm/2.2-3.5°/73-117 sec.
Hatt *et al*. [54]	2015	-	LOOCV:577 images	successful rate/DE	99.8%/0.19 mm
Pourtaherian *et al*. [57]	2015	-	LOOCV:12 volumes	Precision/Recall	0.32/0.75
Mathiassen *et al*. [66]	2016	-	S:512/-/1390 images	95th percentile of DE/t	85% improvement/35.4 FPS
Pourtaherian *et al*. [58]	2017	$\sim 180 \times 190 \times 206$	LOOCV:9 volumes	EE/OE/t	0.60-0.68 mm/2.2-3.7°/>120 sec.
Beigi *et al*. [64]	2017	-	S: 10/5/5 videos	OE/DE	2.12°/1.69 mm
Beigi *et al*. [65]	2017	-	S: 18/6/36 videos	success rate/OE/TE	100%/1.28°/0.82 mm
Mwikirize *et al*. [62]	2017		S:40/-/40 volumes	Precision/Recall/t/TE	0.88/0.98/3.5 sec./0.44 mm
Zanjani *et al*. [51]	2018	$\sim 184 \times 249 \times 203$	S:8/-/2 $\times$3 volumes	DSC improvement	10-20%
Mwikirize *et al*. [77]	2018	-	10-folds: 2500 images	DSC/OE/TE/t	0.99/0.82°/0.23 mm/0.58 sec.
Yang *et al*. [60]	2018	$\sim 152 \times 163 \times 110$	LOOCV:20/10/12 volumes	DSC	0.579-0.744
Younes *et al*. [63]	2018	$765 \times 575 \times 65$	9 volumes for EM	TE/OE/t	4.2 mm/6°/1.25 sec. per needle
Pourtaherian *et al*. [81]	2018	$452 \times 280 \times 292$	5-CV: 20 volumes	visibility	improved
Pourtaherian *et al*. [74]	2018	$\sim 300 \times 230 \times 230$	5-CV: 20/20 volumes	DSC/TE	80-84%/<0.7 mm
Yang *et al*. [61]	2019	$\sim 152 \times 163 \times 110$	LOOCV:10/10/12/8 volumes	DSC/TE/t	0.52-0.83/1.9-3.0 mm/$\sim$450 sec.
Yang *et al*. [70]	2019	$\sim 150 \times 170 \times 151$	3-CV: 65 volumes	DSC/EE/t	0.54/2.07 mm/10 sec.
Yang *et al*. [82]	2019	$\sim 157 \times 160 \times 150$	S:62/-/30 volumes	Precision/Recall/t	0.597/0.686/1.1 sec.
Yang *et al*. [83]	2019	$128 \times 128 \times 128$	3-CV: 25 volumes	DSC/EE	0.577/1.8 mm
Mwikirize *et al*. [89]	2019	-	S:5000/1000/700 images	TE/t	0.72 mm/0.094 sec.
Mwikirize *et al*. [88]	2019	256×256	S:7000/-/500 images	TE/t	0.55 mm/67 FPS
Arif *et al*. [87]	2019	$192 \times 256 \times 128$	2-CV: 149 volumes	DE/OE/t	1 mm/ 2°/3-5 FPS
Yang *et al*. [85]	2019	$\sim 157 \times 160 \times 150$	S:62/30 and 3-folds:18 volumes	DSC	0.658-0.696
Min *et al*. [76]	2020	$376 \times 92 \times 88$	2-CV: 8 volumes	DSC/t	0.673/2.2 sec.
Rodgers *et al*. [80]	2020	-	S:210/-/52 images	DE/OE	0.27 mm/0.5°
Zhang *et al*. [68], [67]	2020	$1024 \times 768 \times N$	S:70/21	detection rate/DE	95%/1.01 mm
Zhang *et al*. [86]	2020	$\sim 1024 \times 768 \times N$	5-CV: 23 patients	shaft error/TE	0.290 mm/0.442 mm
Zhang *et al*. [90]	2020	$\sim 1024 \times 768 \times N$	5-CV: 10 patients	shaft error/TE	0.15 mm/0.44 mm
Lee *et al*. [79]	2020	440×500	S: 794/-/202 images	DSC/OE	0.567/13.3°

describe the instrument pose in the space by a parametrical representation, which is used by other papers in succeeding years. However, the lack of studying segmentation makes this approach less stable when the US images include complex anatomical structures or noise.

Beigi *et al.* (2016) [48]: This work is an example of different image modalities than the above static images. Beigi *et al.* proposed to exploit the time-domain information based on 2D B-mode images, which detected an invisible needle by considering spatiotemporal feature correlations. These approach is totally different than the above segmentation-modeling strategy, so that it is difficult to fairly compare them. Moreover, due to limited hardware and datasets for 3D B-mode imaging, this method was not studied in 3D+time US images in later literature. Nevertheless, this approach should be useful when a larger amount of medical datasets become available in the future.

Pourtaherian *et al.* (2017) [58]: With preliminary studies of segmentation-modeling, this paper performed extensive studies to detect a needle in 3D B-mode images, by applying orientation-invariant Gabor features with an LSVM classifier. With the proposed more discriminating features, the instrument segmentation performance is higher than with previous studies. However, with this complex feature extraction step with the exhaustively voxel-based classification, the total prediction time for each volume is drastically increased from several seconds to several minutes. This limitation hampers the real-time clinical usage for US-guided intervention therapies.

Yang *et al.* (2019) [85]: As the fast development of deep learning with superior performance than conventional handcrafted feature design, this paper employed a state-of-the-art U-Net structure for semantic segmentation, which obtained better segmentation results and resulted into accurate localization for cardiac catheter detection. This paper is one of the preliminary works for 3D US-guided cardiac catheterization, since it is still not as mature as the needle intervention. Nevertheless, the large image size, limited training datasets with laborious expensive annotations and non-real-time performance hamper the real clinical usage for this method, even though it was validated on valuable clinical datasets.

Zhang *et al.* (2020) [68]: Besides the above deep learning or handcrafted feature design for single instrument detection, this paper is an important example to detect multiple needles in 3D US by employing dictionary learning. The instruments were detected by an unsupervised learning approach without effort for data annotation. This paper showed a different solution for conventional segmentation-modeling approaches. Nevertheless, it requires the image domains to have similar appearance and anatomical structures, which limits the flexibility of the method when US images contain different anatomical structures, such as different heart chambers for cardiac catheterization.

With the above summaries, there are remaining some challenges and limitations for this area, despite the current methods obtain satisfactory results. We discuss these challenges below.

• When considering the commonly used segmentation modeling in both non-data-driven and data-driven methods, it is straightforward method without complex pre-/post-processing when compared to the state-of-the-art methods in computer vision. A successful segmentation indeed leads to a better model fitting and detection results, but this approach was only validated on limited datasets, which cannot ensure a generalization and robustness for real clinical applications. For example, different recording settings and recording US equipment can lead to different ultrasound appearance and image quality, which may challenge the robustness and generalization of the algorithm. As a result, a proper pre-processing, such as domain adaptation or image normalization, is required for future work, which is still rarely studied during past years. Moreover, a better post-processing instead of a straightforward RANSAC modeling should be applied to avoid time-consuming processing. Meanwhile, this straightforward fitting method can have limitations when the instrument has a spatial shape different from a curved tube, such as an instrument with ball shape or circular shape for cardiology.

• As for data-driven methods, especially deep learning-based methods, algorithms are tend to learn the information from a large amount of annotated images, which is however difficult to realize in current clinical practice. Because there are many difficulties to record the clinical dataset during an operation, a more matured data collection protocol is required. Moreover, even when it is allowed to collect patient datasets, it is still an expensive solution to train a satisfying model for clinical usage, because of complex network design and the requirement of huge data collection. To overcome these limitations, several solutions may be considered. (1) A simpler and task-specified network design should be employed to decrease overfitting and total detection time to support real-time application. (2) Domain adaptation may be a solution to address the dataset limitation, which can train a network based on *in-vitro*/*ex-vivo* datasets and adapt it to *in-vivo* dataset domain with limited clinical data. This approach is more cost-effective than directly train the model from patient data, as it requires less support from surgeons.

• Considering the existing literature summary, most researchers are focusing on the easiest tasks for US-guided intervention, i.e. needle-based anesthesia or biopsy, since the datasets are easier to obtain than with prostate brachytherapy and cardiac catheterizations. This trend also reflects that these interventions are more mature and widely accepted by hospitals because of their ease of use and lower costs. Nevertheless, researchers should consider more cooperation with hospital and industry to develop instrument detection algorithms for different clinical practice and better surgical outcomes.

• As discussed in this paper, the current literature is focusing on improving detection accuracy rather than time efficiency, which are however both important for US-guided intervention therapies. Moreover, it is unclear for clinical application how accurate the detection should be for broad acceptance by surgeons, e.g. whether 1-mm detection accuracy or 0.7 DSC segmentation accuracy is sufficient for clinical experts. In the future, a more comprehensive study should be performed under the cooperation with surgeons, which should validate the importance and different value settings for detection accuracy and detection efficiency.

• The most recent solutions concentrate on static US images,

which implies to perform detection algorithms on a frame-by-frame basis in the real application. However, in clinical scenarios, real-time US imaging is captured as a video sequence in both 2D and 3D US formats. To better exploit temporal information, 3D+time data should be exploited in future, which is however limited by current software and hardware implementations.

• It is important to have benchmarking datasets for different clinical applications, such as biopsy, anesthesia and cardiac interventions, which enable a fair comparison of the performance and robustness of the methods. This can lead to a better diversity for dedicated solutions and their broad acceptance in the medical community.

VI. Conclusion

In this review paper, we have thoroughly reviewed the technologies for medical instrument detection in US images to guide various types of interventions. We have observed that segmentation modeling is used in the majority of the methods in this area, which also follows the development of computer vision in recent years. Moreover, different clinical applications are reviewed, which shows that the most popular topics are related to needle-based interventions, since it has been widely used in current clinical practice, ranging from anesthesia to biopsy. In contrast, catheter-based methods are not widely considered, because it is still guided by X-ray or other image modalities. The presented overview tables for this domain clearly show that the technology trend follows that of the computer vision area, which leads to the conclusion that all image-based method for interventions have similar limitations from different non-medical-related areas. Summarizing, this paper points out the advantages of recent papers for US-guided interventions, but also bring the weaknesses of these methods and US-based processing to the foreground. In this sense, this review encourages other researchers to explore the less popular areas of US-guided intervention therapy.

Appendix A
Literature Selection

Conference proceedings was searched for IPCAI, MICCAI, SPIE Medical Imaging, IEEE ISBI, IEEE IUS, IEEE ICIP and IEEE EMBC based on the title and abstract of papers. Moreover, related major journal articles are also searched for IEEE TMI, IEEE TBME, IEEE JBHI, IEEE TUFFC, MedIA and IJCARS. With specified searching input, we considered search string as '(Needle OR Catheter OR Instrument) AND (Detection OR Segmentation OR Localization) AND Ultrasound'. We went over all the searched papers by title and abstract to make sure the content is correct. If there were still some misleading, we went to the main content of the paper to make a decision. We have excluded the paper with external device as the detection support, such as EM sensor or robotics (paper was only selected when there were some image processing solutions applied for instrument detection). We also checked references of the papers to confirm the key publications were not missing, as some of papers might be published in other journals or conferences.

References

[1] B. R. Douglas, J. W. Charboneau, and C. C. Reading, "Ultrasound-guided intervention: expanding horizons," *Radiologic Clinics of North America*, vol. 39, no. 3, pp. 415–428, 2001.

[2] I. M. Germano, *Advanced techniques in image-guided brain and spine surgery*. Thieme Medical Publishers, Incorporated, 2002.

[3] T. M. Peters, "Image-guidance for surgical procedures," *Physics in Medicine & Biology*, vol. 51, no. 14, p. R505, 2006.

[4] K. Cleary and T. M. Peters, "Image-guided interventions: technology review and clinical applications," *Annual review of biomedical engineering*, vol. 12, pp. 119–142, 2010.

[5] K. A. Scanlan, P. A. Propeck, and F. T. Lee Jr, "Invasive procedures in the female pelvis: value of transabdominal, endovaginal, and endorectal us guidance," *Radiographics*, vol. 21, no. 2, pp. 491–506, 2001.

[6] T. Hatada, H. Ishii, S. Ichii, K. Okada, Y. Fujiwara, and T. Yamamura, "Diagnostic value of ultrasound-guided fine-needle aspiration biopsy, core-needle biopsy, and evaluation of combined use in the diagnosis of breast lesions," *Journal of the American College of Surgeons*, vol. 190, no. 3, pp. 299–303, 2000.

[7] D. E. Coplen, G. L. Andrile, J. J. Yuan, and W. J. Catalona, "The ability of systematic transrectal ultrasound guided biopsy to detect prostate cancer in men with the clinical diagnosis of benign prostatic hyperplasia," *The Journal of urology*, vol. 146, no. 1, pp. 75–77, 1991.

[8] M. J. Barrington and R. Kluger, "Ultrasound guidance reduces the risk of local anesthetic systemic toxicity following peripheral nerve blockade," 2013.

[9] D. H. Sheafor, E. K. Paulson, C. M. Simmons, D. M. DeLong, and R. C. Nelson, "Abdominal percutaneous interventional procedures: comparison of ct and us guidance." *Radiology*, vol. 207, no. 3, pp. 705–710, 1998.

[10] J. Machi, S. Uchida, K. Sumida, W. M. Limm, S. A. Hundahl, A. J. Oishi, N. L. Furumoto, and R. H. Oishi, "Ultrasound-guided radiofrequency thermal ablation of liver tumors: percutaneous, laparoscopic, and open surgical approaches," *Journal of Gastrointestinal Surgery*, vol. 5, no. 5, pp. 477–489, 2001.

[11] D. Oepkes, R. Devlieger, E. Lopriore, and F. Klumper, "Successful ultrasound-guided laser treatment of fetal hydrops caused by pulmonary sequestration," *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, vol. 29, no. 4, pp. 457–459, 2007.

[12] W. Xia *et al.*, "In-plane ultrasonic needle tracking using a fiber-optic hydrophone," *Medical Physics*, vol. 42, no. 10, pp. 5983–5991, 2015.

[13] J. Krücker, S. Xu, N. Glossop, A. Viswanathan, J. Borgert, H. Schulz, and B. J. Wood, "Electromagnetic tracking for thermal ablation and biopsy guidance: clinical evaluation of spatial accuracy," *Journal of Vascular and Interventional Radiology*, vol. 18, no. 9, pp. 1141–1150, 2007.

[14] C. Nadeau *et al.*, "Intensity-based visual servoing for instrument and tissue tracking in 3d ultrasound volumes," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 1, pp. 367–371, 2014.

[15] M. Ding, H. N. Cardinal, W. Guan, and A. Fenster, "Automatic needle segmentation in 3d ultrasound images," in *Medical Imaging 2002: Visualization, Image-Guided Procedures, and Display*, vol. 4681. International Society for Optics and Photonics, 2002, pp. 65–76.

[16] M. Ding and A. Fenster, "Projection-based needle segmentation in 3d ultrasound images," *Computer Aided Surgery*, vol. 9, no. 5, pp. 193–201, 2004.

[17] S. H. Okazawa, R. Ebrahimi, J. Chuang, R. N. Rohling, and S. E. Salcudean, "Methods for segmenting curved needles in ultrasound images," *Medical image analysis*, vol. 10, no. 3, pp. 330–342, 2006.

[18] H. Zhou, W. Qiu, M. Ding, and S. Zhang, "Automatic needle segmentation in 3d ultrasound images using 3d hough transform," in *MIPPR 2007: Medical Imaging, Parallel Processing of Images, and Optimization Techniques*, vol. 6789. International Society for Optics and Photonics, 2007, p. 67890R.

[19] ——, "Automatic needle segmentation in 3d ultrasound images using 3d improved hough transform," in *Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*, vol. 6918. International Society for Optics and Photonics, 2008, p. 691821.

[20] W. Qiu, M. Yuchi, M. Ding, D. Tessier, and A. Fenster, "Needle segmentation using 3d hough transform in 3d trus guided prostate transperineal therapy," *Medical physics*, vol. 40, no. 4, p. 042902, 2013.

[21] M. Barva, M. Uhercik, J.-M. Mari, J. Kybic, J.-R. Duhamel, H. Liebgott, V. Hlavác, and C. Cachard, "Parallel integral projection transform for straight electrode localization in 3-d ultrasound images," *ieee transac-*

tions on ultrasonics, ferroelectrics, and frequency control, vol. 55, no. 7, pp. 1559–1569, 2008.

[22] M. Aboofazeli, P. Abolmaesumi, P. Mousavi, and G. Fichtinger, "A new scheme for curved needle segmentation in three-dimensional ultrasound images," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2009, pp. 1067–1070.

[23] P. Beigi and R. Rohling, "Needle localization using a moving stylet/catheter in ultrasound-guided regional anesthesia: a feasibility study," in *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9036. International Society for Optics and Photonics, 2014, p. 90362Q.

[24] M. I. Daoud, A. Shtaiyat, A. R. Zayadeen, and R. Alazrai, "Accurate needle localization using two-dimensional power doppler and b-mode ultrasound image analyses: A feasibility study," *Sensors*, vol. 18, no. 10, p. 3475, 2018.

[25] Y. Zhao, Y. Shen, A. Bernard, C. Cachard, and H. Liebgott, "Evaluation and comparison of current biopsy needle localization and tracking methods using 3d ultrasound," *Ultrasonics*, vol. 73, pp. 206–220, 2017.

[26] P. M. Novotny, J. W. Cannon, and R. D. Howe, "Tool localization in 3d ultrasound images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2003, pp. 969–970.

[27] M. G. Linguraru and R. D. Howe, "Texture-based instrument segmentation in 3d ultrasound images," in *Medical Imaging 2006: Image Processing*, vol. 6144. International Society for Optics and Photonics, 2006, p. 61443J.

[28] S. Zhao, W. Qiu, Y. Ming, and M. Ding, "Needle segmentation in 3d ultrasound images based on phase grouping," in *MIPPR 2009: Medical Imaging, Parallel Processing of Images, and Optimization Techniques*, vol. 7497. International Society for Optics and Photonics, 2009, p. 74971L.

[29] I. McSweeney, B. Murphy, and W. M. Wright, "Estimation of needle tip location using ultrasound image processing and hypoechoic markers," in *2014 IEEE International Ultrasonics Symposium*. IEEE, 2014, pp. 1876–1879.

[30] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International conference on medical image computing and computer-assisted intervention*. Springer, 1998, pp. 130–137.

[31] H. Ren and P. E. Dupont, "Tubular structure enhancement for surgical instrument detection in 3d ultrasound," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 7203–7206.

[32] O. Mohareri, M. Ramezani, T. K. Adebar, P. Abolmaesumi, and S. E. Salcudean, "Automatic localization of the da vinci surgical instrument tips in 3-d transrectal ultrasound," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 9, pp. 2663–2672, 2013.

[33] Y. Zhao, C. Cachard, and H. Liebgott, "Automatic needle detection and tracking in 3d ultrasound using an roi-based ransac and kalman method," *Ultrasonic imaging*, vol. 35, no. 4, pp. 283–306, 2013.

[34] L. Malekian, H. A. Talebi, and F. Towhidkhah, "A noise adaptive method for needle localization in 3d ultrasound images," in *2014 Iranian Conference on Intelligent Systems (ICIS)*. IEEE, 2014, pp. 1–5.

[35] N. Agarwal, A. K. Yadav, A. Gupta, and M. F. Orlando, "Real-time needle tip localization in 2d ultrasound images using kalman filter," in *2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2019, pp. 1008–1012.

[36] Y. Zhao, C. Cachard, and H. Liebgott, "A new automatically biopsy needle tracking method using 3d ultrasound," in *2013 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2013, pp. 844–847.

[37] A. Pourtaherian, N. Mihajlovic, S. Zinger, H. H. Korsten, P. H. de With, J. Huang, and G. C. Ng, "Automated in-plane visualization of steep needles from 3d ultrasound data volumes," in *2016 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2016, pp. 1–4.

[38] K. Cao, D. Mills, and K. A. Patwardhan, "Automated catheter detection in volumetric ultrasound," in *2013 IEEE 10th International Symposium on Biomedical Imaging*. IEEE, 2013, pp. 37–40.

[39] M. Kaya and O. Bebek, "Gabor filter based localization of needles in ultrasound guided robotic interventions," in *2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings*. IEEE, 2014, pp. 112–117.

[40] ——, "Needle localization using gabor filtering in 2d ultrasound images," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 4881–4886.

[41] M. Kaya, E. Senel, A. Ahmad, O. Orhan, and O. Bebek, "Real-time needle tip localization in 2d ultrasound images for robotic biopsies," in *2015 International Conference on Advanced Robotics (ICAR)*. IEEE, 2015, pp. 47–52.

[42] I. Hacihaliloglu, P. Beigi, G. Ng, R. N. Rohling, S. Salcudean, and P. Abolmaesumi, "Projection-based phase features for localization of a needle tip in 2d curvilinear ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 347–354.

[43] C. Mwikirize, J. L. Nosher, and I. Hacihaliloglu, "Enhancement of needle tip and shaft from 2d ultrasound using signal transmission maps," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 362–369.

[44] P. H. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Computer vision and image understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[45] P. Beigi, S. E. Salcudean, R. Rohling, and G. C. Ng, "Automatic detection of a hand-held needle in ultrasound via phased-based analysis of the tremor motion," in *Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9786. International Society for Optics and Photonics, 2016, p. 97860I.

[46] M. Kaya, E. Senel, A. Ahmad, and O. Bebek, "Visual tracking of biopsy needles in 2d ultrasound images," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 4386–4391.

[47] P. Beigi, T. Salcudean, R. Rohling, V. A. Lessoway, and G. C. Ng, "Needle detection in ultrasound using the spectral properties of the displacement field: a feasibility study," in *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9415. International Society for Optics and Photonics, 2015, p. 94150U.

[48] P. Beigi, R. Rohling, S. E. Salcudean, and G. C. Ng, "Spectral analysis of the tremor motion for needle detection in curvilinear ultrasound via spatiotemporal linear sampling," *International journal of computer assisted radiology and surgery*, vol. 11, no. 6, pp. 1183–1192, 2016.

[49] M. I. Daoud, A.-L. Alshalalfah, O. A. Mohamed, and R. Alazrai, "A hybrid camera-and ultrasound-based approach for needle localization and tracking using a 3d motorized curvilinear ultrasound probe," *Medical image analysis*, vol. 50, pp. 145–166, 2018.

[50] J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane, "Automated design of both the topology and sizing of analog electrical circuits using genetic programming," in *Artificial Intelligence in Design'96*. Springer, 1996, pp. 151–170.

[51] F. G. Zanjani, A. Pourtaherian, X. Tang, S. Zinger, N. Mihajlovic, G. C. Ng, H. H. Korsten *et al.*, "Coherent needle detection in ultrasound volumes using 3d conditional random fields," in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 10576. International Society for Optics and Photonics, 2018, p. 105760W.

[52] D. Krefting, B. Haupt, T. Tolxdorff, C. Kempkensteffen, and K. Miller, "Segmentation of prostate biopsy needles in transrectal ultrasound images," in *Medical Imaging 2007: Image Processing*, vol. 6512. International Society for Optics and Photonics, 2007, p. 65122Y.

[53] M. Uherčík, J. Kybic, Y. Zhao, C. Cachard, and H. Liebgott, "Line filtering for surgical tool localization in 3d ultrasound images," *Computers in biology and medicine*, vol. 43, no. 12, pp. 2036–2045, 2013.

[54] C. R. Hatt, G. Ng, and V. Parthasarathy, "Enhanced needle localization in ultrasound using beam steering and learning-based segmentation," *Computerized Medical Imaging and Graphics*, vol. 41, pp. 46–54, 2015.

[55] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[56] A. Pourtaherian, S. Zinger, H. H. Korsten, N. Mihajlovic *et al.*, "Benchmarking of state-of-the-art needle detection algorithms in 3d ultrasound data volumes," in *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9415. International Society for Optics and Photonics, 2015, p. 94152B.

[57] A. Pourtaherian, S. Zinger, N. Mihajlovic, J. Huang, G. C. Ng, H. H. Korsten *et al.*, "Multi-resolution gabor wavelet feature extraction for needle detection in 3d ultrasound," in *Eighth International Conference on Machine Vision (ICMV 2015)*, vol. 9875. International Society for Optics and Photonics, 2015, p. 987513.

[58] A. Pourtaherian, H. J. Scholten, L. Kusters, S. Zinger, N. Mihajlovic, A. F. Kolen, F. Zuo, G. C. Ng, H. H. Korsten, and P. H. de With, "Medical instrument detection in 3-dimensional ultrasound data volumes," *IEEE transactions on medical imaging*, vol. 36, no. 8, pp. 1664–1675, 2017.

[59] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[60] H. Yang, A. Pourtaherian, C. Shan, A. F. Kolen *et al.*, "Feature study on catheter detection in three-dimensional ultrasound," in *Medical Imaging*

2018: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 10576. International Society for Optics and Photonics, 2018, p. 105760V.

[61] H. Yang, C. Shan, A. Pourtaherian, A. F. Kolen *et al.*, "Catheter segmentation in three-dimensional ultrasound images by feature fusion and model fitting," *Journal of Medical Imaging*, vol. 6, no. 1, p. 015001, 2019.

[62] C. Mwikirize, J. L. Nosher, and I. Hacihaliloglu, "Local phase-based learning for needle detection and localization in 3d ultrasound," in *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures*. Springer, 2017, pp. 108–115.

[63] H. Younes, S. Voros, and J. Troccaz, "Automatic needle localization in 3d ultrasound images for brachytherapy," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1203–1207.

[64] P. Beigi, R. Rohling, T. Salcudean, V. A. Lessoway, and G. C. Ng, "Detection of an invisible needle in ultrasound using a probabilistic svm and time-domain features," *Ultrasonics*, vol. 78, pp. 18–22, 2017.

[65] P. Beigi, R. Rohling, S. E. Salcudean, and G. C. Ng, "Casper: computer-aided segmentation of imperceptible motion—a learning-based tracking of an invisible needle in ultrasound," *International journal of computer assisted radiology and surgery*, vol. 12, no. 11, pp. 1857–1866, 2017.

[66] K. Mathiassen, D. Dall'Alba, R. Muradore, P. Fiorini, and O. J. Elle, "Robust real-time needle tracking in 2-d ultrasound images using statistical filtering," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 966–978, 2016.

[67] Y. Zhang, X. He, Z. Tian, J. Jeong, Y. Lei, T. Wang, Q. Zeng, A. B. Jani, W. Curran, P. Patel *et al.*, "Multi-needle detection in 3d ultrasound images with sparse dictionary learning," in *Medical Imaging 2020: Ultrasonic Imaging and Tomography*, vol. 11319. International Society for Optics and Photonics, 2020, p. 113190I.

[68] Y. Zhang, X. He, Z. Tian, J. J. Jeong, Y. Lei, T. Wang, Q. Zeng, A. B. Jani, W. J. Curran, P. Patel *et al.*, "Multi-needle detection in 3d ultrasound images using unsupervised order-graph regularized sparse dictionary learning," *IEEE Transactions on Medical Imaging*, 2020.

[69] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[70] H. Yang, C. Shan, A. F. Kolen *et al.*, "Catheter localization in 3d ultrasound using voxel-of-interest-based convnets for cardiac intervention," *International journal of computer assisted radiology and surgery*, vol. 14, no. 6, pp. 1069–1077, 2019.

[71] A. A. Geraldes and T. S. Rocha, "A neural network approach for flexible needle tracking in ultrasound images using kalman filter," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 2014, pp. 70–75.

[72] T. S. Rocha and A. A. Geraldes, "Flexible needles detection in ultrasound images using a multi-layer perceptron network," in *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC)*. IEEE, 2014, pp. 1–5.

[73] A. Pourtaherian, F. G. Zanjani, S. Zinger, N. Mihajlovic, G. Ng, H. Korsten *et al.*, "Improving needle detection in 3d ultrasound using orthogonal-plane convolutional networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 610–618.

[74] A. Pourtaherian, F. G. Zanjani, S. Zinger, N. Mihajlovic, G. C. Ng, H. H. Korsten *et al.*, "Robust and semantic needle detection in 3d ultrasound using orthogonal-plane convolutional neural networks," *International journal of computer assisted radiology and surgery*, vol. 13, no. 9, pp. 1321–1333, 2018.

[75] H. Yang, C. Shan, A. F. Kolen, and P. H. de With, "Catheter detection in 3d ultrasound using triplanar-based convolutional neural networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 371–375.

[76] L. Min, H. Yang, C. Shan, A. F. Kolen *et al.*, "Feasibility study of catheter segmentation in 3d frustum ultrasounds by dcnn," in *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11315. International Society for Optics and Photonics, 2020, p. 1131521.

[77] C. Mwikirize, J. L. Nosher, and I. Hacihaliloglu, "Convolution neural networks for real-time needle detection and localization in 2d ultrasound," *International journal of computer assisted radiology and surgery*, vol. 13, no. 5, pp. 647–657, 2018.

[78] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[79] J. Y. Lee, M. Islam, J. R. Woh, T. M. Washeem, L. Y. C. Ngoh, W. K. Wong, and H. Ren, "Ultrasound needle segmentation and trajectory prediction using excitation network," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 3, pp. 437–443, 2020.

[80] J. R. Rodgers, D. J. Gillies, W. T. Hrinivich, I. Gyackov, and A. Fenster, "Automatic needle localization in intraoperative 3d transvaginal ultrasound images for high-dose-rate interstitial gynecologic brachytherapy," in *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 11315. International Society for Optics and Photonics, 2020, p. 113150K.

[81] A. Pourtaherian, N. Mihajlovic, F. GhazvinianZanjani, S. Zinger, G. C. Ng, H. H. Korstcn, and P. H. De With, "Localization of partially visible needles in 3d ultrasound using dilated cnns," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–4.

[82] H. Yang, C. Shan, A. F. Kolen, and P. H. de With, "Efficient catheter segmentation in 3d cardiac ultrasound using slice-based fcn with deep supervision and f-score loss," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 260–264.

[83] H. Yang, C. Shan, A. F. Kolen, and H. de With Peter, "Improving catheter segmentation & localization in 3d cardiac ultrasound using direction-fused fcn," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1122–1126.

[84] H. Yang, C. Shan, A. F. Kolen, and P. H. de With, "Automated catheter localization in volumetric ultrasound using 3d patch-wise u-net with focal loss," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1346–1350.

[85] H. Yang, C. Shan, T. Tan, A. F. Kolen *et al.*, "Transferring from ex-vivo to in-vivo: Instrument localization in 3d cardiac ultrasound using pyramid-unet with hybrid loss," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 263–271.

[86] Y. Zhang, Y. Lei, R. L. Qiu, T. Wang, H. Wang, A. B. Jani, W. J. Curran, P. Patel, T. Liu, and X. Yang, "Multi-needle localization with attention u-net in us-guided hdr prostate brachytherapy," *Medical Physics*, 2020.

[87] M. Arif, A. Moelker, and T. van Walsum, "Automatic needle detection and real-time bi-planar needle visualization during 3d ultrasound scanning of the liver," *Medical image analysis*, vol. 53, pp. 104–110, 2019.

[88] C. Mwikirize, J. L. Nosher, and I. Hacihaliloglu, "Single shot needle tip localization in 2d ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 637–645.

[89] ——, "Learning needle tip localization from digital subtraction in 2d ultrasound," *International journal of computer assisted radiology and surgery*, vol. 14, no. 6, pp. 1017–1026, 2019.

[90] Y. Zhang, J. Harms, Y. Lei, T. Wang, T. Liu, A. B. Jani, W. J. Curran, P. Patel, and X. Yang, "Weakly supervised multi-needle detection in 3d ultrasound images with bidirectional convolutional sparse coding," in *Medical Imaging 2020: Ultrasonic Imaging and Tomography*, vol. 11319. International Society for Optics and Photonics, 2020, p. 1131914.

[91] A. Pourtaherian, "Robust needle detection and visualization for 3d ultrasound image-guided interventions," Ph.D. dissertation, Department of Electrical Engineering, 9 2018, proefschrift.

[92] S. A. Patel, K. Pierko, and R. Franco-Sadud, "Ultrasound-guided bedside core needle biopsy: A hospitalist procedure team's experience," *Cureus*, vol. 11, no. 1, 2019.

[93] S. Banerjee, T. Kataria, D. Gupta, S. Goyal, S. S. Bisht, T. Basu, and A. Abhishek, "Use of ultrasound in image-guided high-dose-rate brachytherapy: enumerations and arguments," *Journal of contemporary brachytherapy*, vol. 9, no. 2, p. 146, 2017.

[94] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10TH NORDICBALTIC CONFERENCE ON BIOMEDICAL IMAGING, VOL. 4, SUPPLEMENT 1, PART 1: 351–353*. Citeseer, 1996.