Keypoint Localization Based on Convolutional Neural Network for Robotic Implantation of Flexible Micro-Electrodes

Wenliang Liang, Fangbo Qin, Xinyong Han, Dapeng Zhang*

***Abstract*— Visual localization of micro flexible electrode and implant needle is an important task for robotic flexible electrode implantation. Magnification switch, occlusion, defocus, illumination changes in microscopic imaging produce challenges for this task. We propose the Keypoint Localization and Angle Estimation Network (KLAE-Net) based on convolutional neural networks. KLAE-Net has two branches: the keypoint localization branch for obtaining the coordinates of electrode and needle in image space; the angle estimation branch for monitoring the inclination of needle. Attention mechanism and deformable convolution are used to improve the model's performance. For training and evaluation under the flexible electrode implantation task, we construct a novel dataset containing 1000 images covering various conditions. An image Jacobian matrix based alignment control method is designed, to realize the robotic alignment between needle and electrode. A series of experiments are conducted with the dataset and an implantation robot system.**

I. INTRODUCTION

Brain-machine interface (BMI), as a multidisciplinary technology of neuroscience, electronics, artificial intelligence and robotics, has attracted wide attention of researchers in recent years [1]. With BMI, a human can control external devices by converting brain neuron activity into specific instructions. BMI devices includes invasive and non-invasive ones. Non-invasive BMI directly records scalp electroencephalogram (EEG) without trauma and surgical risk, but the signal information is limited, and there are bottlenecks in real-time and accuracy [2]. Invasive BMI requires implanting electrodes into neural tissue inside the skull to collect brain signals. It can record electrical signals at the neuron level with large amount of information and has better performance in real-time and accuracy.

The traditional rigid electrodes have a significant mechanical mismatch with neural tissue, which induces immune response and callus, and then lead to the deterioration or disappearance of signal [3]. In order to reduce the adverse effects, the size and mechanical stiffness of implanted electrode should be reduced. Flexible electrodes with micrometer-level diameters have good biocompatibility and reliable signal-collection quality [4]. However, due to the small Young's modulus of the flexible material, it is easy to deform and difficult to manipulate in the implantation process. A variety of auxiliary methods, including temporary changes in flexible electrode stiffness [5], removable auxiliary implants [6], and degradable templates [7] have been proposed to improve the rigidity of implantable flexible electrodes and successfully achieve precise implantation of flexible neural electrodes. In order to make electrode implantation more precise, efficient and reliable, minimally invasive implant robot acts as an important role.

The accurate localization of flexible electrode and implant tools relies on microscopic cameras. By detecting and localizing the keypoints on electrodes and tools, the image keypoints can be provided to guide the control of tools to manipulate flexible electrode. The existing keypoint localization methods can be summarized into two categories. The first category uses the traditional computer vision method [8-10]. These methods detect or track the instrument's parts by using handcrafted features and extract low-level visual features around keypoint to learn the appearance templates. The second category is based on deep learning [11]. Recently, with the extensive application of deep learning methods in surgical vision, some methods of using deep CNNs to localize keypoint of the surgical instrument have emerged [12-15]. Since the deep learning methods can extract multi-level and multi-scale information, their localization performances are greatly improved compared with those of the traditional methods. SR-Net [14] extends the U-Net segmentation model to realize keypoint localization. G-RMI [12] uses fully convolutional ResNet to predict activation heatmaps and offsets for each keypoint. G-RMI detects bounding boxes that contain objects, then estimate the keypoints that each proposal bounding box contain. Hourglass [13] applies intermediate supervision to repeated down- and up-sampling processes for keypoint localization. Mask R-CNN [15] is one of the most popular frameworks for instance segmentation, which can accomplish instance detection and the segmentation for each instance a single model. Mask R-CNN can be easily expanded for keypoint localization by customizing the output.

In this paper, we have two main motivations. First, the visual localization of flexible electrode and implant tool is required for robotic implantation. Considering the varying

This work is supported by National Key Research and Development Program of China (2021ZD0200402), and National Natural Science Foundation of China (62103413, 62073317).

W. Liang, F. Qin, X. Han and D. Zhang are with Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. W. Liang is also with School of Artificial Intelligence University of Chinese Academy of Science, Beijing 100049, China. {qinfangbo2013; dapeng.zhang@ia.ac.cn}.

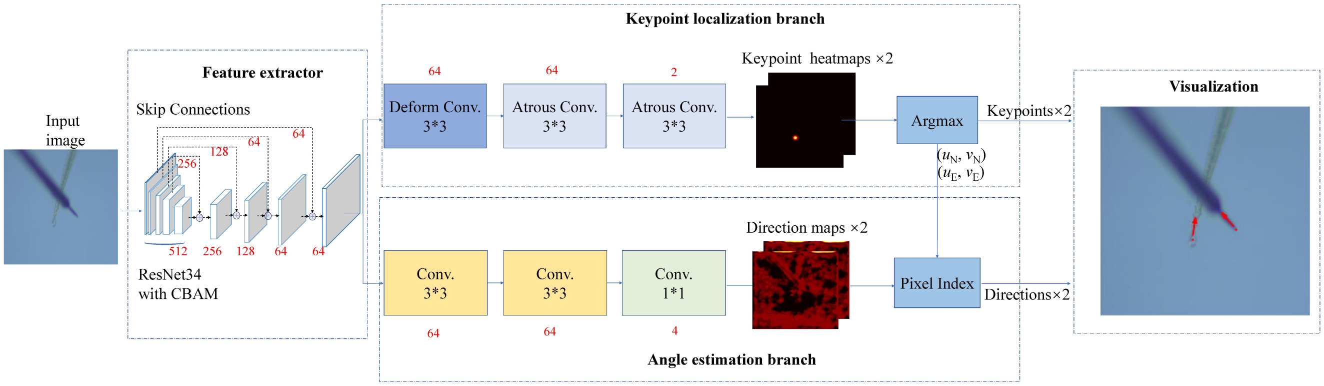

Fig. 1. Model Architecture. The output channel number of layers are labeled by red digits.

magnification, illumination, clearness and the occurring of partial occlusion, we exploit deep CNN to realize the detection and localization of the relevant keypoints. Besides, the inclined angle of electrode and tool should be estimated. Second, guided by the obtained keypoints in real-time images, the implant robot can use an implant needle to hook the ring at the end of electrode. The main contributions of this paper are as follows:

1) We propose the keypoint localization and angle estimation network (KLAE-Net) to realize the automatic pose feature extraction from microscopic image. The keypoint features are used for alignment and the angle features indicate the inclination of implant tool.

2) A novel dataset is collected under the flexible electrode implantation scene, which contains 1000 images and annotations. The images involve various illumination, clearness, magnification and occlusion conditions.

3) Using the implant needle tip and the electrode ring center as the image keypoint features, the alignment control of needle and the electrode is realized with image Jacobian matrix and feedback controller.

II. TASK DESCRIPTION

Flexible electrode implantation workflow: Similar to [20], the implantation robot controls an implant needle and moves the needle tip to hook the ring at the end of electrode, which is guided by microscopic vision. Then a pincher rotates and grips the flexible electrode's body. Then, the flexible electrode is moved along with the implant needle and is implanted into the target brain as planned. Finally, the implant needle is rapidly withdrawn, and the electrode is retained in the brain tissue to collect the signals of the specific target area of the brain, thus realizing the construction of the signal acquisition pathway of the brain-machine interface.

Visual localization task: The binocular microscopic cameras collect the microscopic images of needle implant needle and flexible electrode in real time. Aiming to sense the actual pose of implant needle tip and electrode tip for visual servoing, the precise position of the implant needle tip and the electrode ring center, as well as the angles of the needle axis and electrode body, should be extracted from the microscopic images. Then the relative position between implant needle tip and electrode ring center in Cartesian space can be measured through the pre-calibrated image Jacobian matrix, which is used to guide the visual servoing.

III. METHODS

A. *Model Architecture*

As shown in Fig. 1, the proposed KLAE-Net takes an RGB image I as input. Its outputs include a needle keypoint heatmap H_N, an electrode keypoint heatmap H_E, a needle direction map D_N, and an electrode direction map D_E. The model consists of a feature extractor and two branches for localizing keypoints and estimating angles, respectively.

Feature extractor: The U-shaped feature extractor is formed by a backbone and four decoder blocks. The backbone is implemented with ResNet-34 [16], to extract multi-scale features from the input image. The final output of the backbone is a 512-channel feature map whose size is 1/16 of the input size. Besides, four lower-level feature maps are drawn out from the backbone and used to provide skip-connections to the decoders.

We utilize the lightweight Convolutional Block Attention Module (CBAM) [17] in the backbone, to realize channel and spatial attentions. Each ResNet block is integrated with a CBAM, as shown in Fig. 2. Given an intermediate feature map, CBAM calculates the channel attention map and spatial attention map. Then element-wise multiplication between input feature map and attention maps is used to reweight the features, so that unimportant features are suppressed and relevant features are retained for inference.

In each decoder block, the input feature map is firstly upsampled by 2×, then processed by a 1×1 convolution layer, a batch normalization layer and a ReLU layer. The feature map from skip connection is also processed by a 1×1 convolution layer, a batch normalization layer and a ReLU layer. Then the two processed feature maps are concatenated and inputted to a ResBlock, as shown in Fig. 3. The final output of the feature extractor is a 128-channel feature map F, whose size is the same with the input's.

Keypoint localization branch. The feature map F is processed by three convolution layers to infer the keypoint heatmaps H_N and H_E. The first layer is implemented by deformable convolution (DCN) to realize shape-aware adaptive spatial sampling [19]. The last two layers are

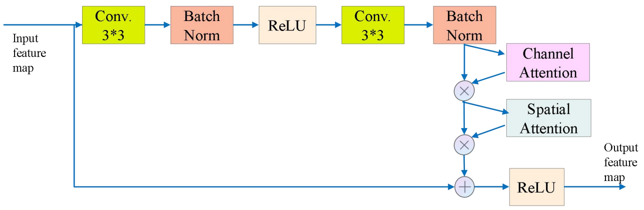

Fig. 2. Architecture of ResBlock with CBAM.

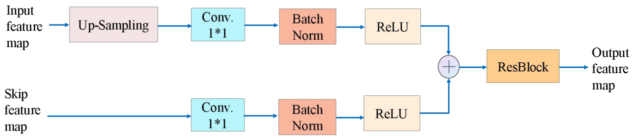

Fig. 3. Architecture of decoder block.

implemented by atrous convolution with the atrous rate of 2 [21]. The final output is a 2-channel map, which is split to the two 1-channel heatmaps H_N and H_E.

Angle estimation branch: The feature map F are also processed by two 3×3 convolution layers and a 1×1 convolution layer to generate pixel-wise direction maps D_N and D_E. Each pixel of a direction map is a 2-D normalized direction vector, indicating the angle in image space.

B. Post-Processing

In the inference stage, the output maps are parsed to the coordinates and angles. Because there are only one needle and one electrode in the microscopic view, the keypoint coordinates $(u_{\text{key}},v_{\text{key}})$ can be obtained from keypoint heatmap H simply by argmax operation, namely,

$$\left(u_{\text{key}},v_{\text{key}}\right)=\underset{u,v}{\arg\max}\,H\left(u,v\right) \tag{1}$$

Afterwards, the direction vector corresponding to the keypoint can be indexed with the aforementioned coordinates, which is converted to scalar angle θ, as given by,

$$\theta=\arctan\frac{D(u_{key},v_{key},1)}{D(u_{key},v_{key},0)} \tag{2}$$

Thus, the implant needle's position and direction are obtained from H_N and D_N, which are expressed as (u_N,v_N,θ_N). The flexible electrode's position and direction are obtained from H_E and D_E, which are expressed as (u_E,v_E,θ_E).

C. Training Loss

The ground-truth of keypoint heatmap H_{GT} is generated by the following equation:

$$H_{\text{GT}}\left(u,v\right)=\exp\left(-\frac{\left(u-u_{GT}\right)^2+\left(v-v_{GT}\right)^2}{\sigma^2}\right) \tag{3}$$

where u_{GT} and v_{GT} are the annotated keypoint's coordinates. The Gaussian standard deviation σ is set as 31 for the 512×512 map size. The keypoint loss is calculated by,

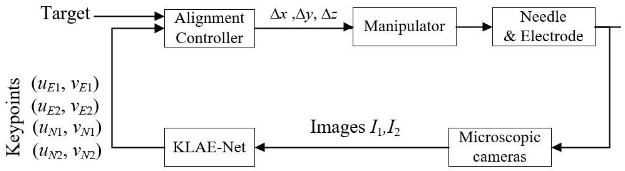

Fig. 4. Block diagram of IBVS feedback control system.

$$\begin{aligned}L_{\text{KP}}=&\frac{1}{N}\sum_{u,v}\left[H_N\left(u,v\right)-H_{NGT}\left(u,v\right)\right]^2\\&+\frac{1}{N}\sum_{u,v}\left[H_E\left(u,v\right)-H_{EGT}\left(u,v\right)\right]^2\end{aligned} \tag{4}$$

where N is the pixel number of the map.

The direction loss is calculated with the keypoint pixels and their neighborhood pixels, as given by,

$$\begin{aligned}L_{\text{D}}=&\frac{1}{N_n+1}\sum_{u,v,i}\left[D_N\left(u,v,i\right)-D_{NGT}\left(u,v,i\right)\right]^2\\&+\frac{1}{N_n+1}\sum_{u,v,i}\left[D_E\left(u,v,i\right)-D_{EGT}\left(u,v,i\right)\right]^2\end{aligned} \tag{5}$$

where N_n is the neighborhood pixel number, whose default value is 8. The total loss is the sum of the above losses,

$$L_{\text{total}}=L_{\text{KP}}+L_{\text{D}} \tag{6}$$

IV. ALIGNMENT CONTROL BASED ON VISUAL SERVOING

Compared to position-based visual servoing (PBVS) that is sensitive to calibration error, image-based visual servoing (IBVS) is utilized in this work due to its insensitivity to errors caused by calibration and feature extraction [22]. IBVS uses the image feature errors between the implanted needle tip and electrode ring center in the binocular images as feedbacks, to generate the 3-D motion of the robotic manipulator.

For microscopic vision servoing, the relationship between the 3-D position error (Δx, Δy, Δz) in Cartesian space and the 2-D position errors (Δu_1, Δv_1) (Δu_2, Δv_2) in two microscopic image spaces is modeled with image Jacobian matrix J, as given by,

$$\begin{bmatrix}\Delta u_1\\\Delta v_1\\\Delta u_2\\\Delta u_2\end{bmatrix}=\begin{bmatrix}J_{11}&J_{12}&J_{13}\\J_{21}&J_{22}&J_{23}\\J_{31}&J_{32}&J_{33}\\J_{41}&J_{42}&J_{43}\end{bmatrix}\begin{bmatrix}\Delta x\\\Delta y\\\Delta z\end{bmatrix}=J\begin{bmatrix}\Delta x\\\Delta y\\\Delta z\end{bmatrix} \tag{7}$$

The image Jacobian matrix J can be calibrated with least square method.

As shown in Fig. 4, the alignment controller takes the position errors between the needle tip and the electrode ring center in the two microscopic images as the inputs. The output is calculated with the pseudo-inverse of the image Jacobian matrix and linear feedback control law, namely,

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = KJ^{\dagger} \begin{bmatrix} u_{E1} - u_{N1} \\ v_{E1} - v_{N1} \\ u_{E2} - u_{N2} \\ v_{E1} - v_{N1} \end{bmatrix} \tag{8}$$

where K is the control gain, whose default value is 0.5. (u_{E1}, v_{E1}) and (u_{E2}, v_{E2}) are the electrode keypoint coordinates in the images of camera 1 and 2, respectively. (u_{N1}, v_{N1}) and (u_{N2}, v_{N2}) are the needle keypoint coordinates in the images of camera 1 and 2, respectively. These image coordinates are all provided by the proposed KLAE-Net. The control target is to reduce the image errors to zero.

V. EXPERIMENTS AND RESULTS

A. *Experiment Platform*

The implantation robotic system mainly consists of a UR5 robot, two motorized precision stages, two Navitar microscopic cameras, and the implantation tools, as shown in Fig. 5. The UR5 robot is used to control the pose of implantation tool in a large range. The precision stage is used to fine-tune the position with micrometer precision. The microscopic cameras can change their magnifications from 0.35× to 2.25× with step motors, whose position can be adjusted by Winner Optical Instruments linear stages. As shown in the right picture in Fig. 5, the implantation tools at the robot's end include an implant needle and a L-shaped pincher. The implant needle is sharp at its end, and the tip diameter is about 10μm. The flexible electrode's width is about 50μm and the ring at its end has a dimeter of ~30μm. Thus, the needle tip is able to move through the electrode ring. The image Jacobian matrices has been calibrated using the least square method.

B. *Training Details and Evaluation Metrics*

We construct the image dataset for flexible electrode implantation. 1000 images are collected with our robot system and manually labeled by experts. 1000 images can be divided into 350 small magnification, 200 medium magnification, 450 large magnification; 200 with occlusion, 800 without occlusion; 150 out of focus, 850 on focus; 250 low light, 500 normal light, 250 strong light. 900 images are used for training and 100 images for evaluation. The size of the original images is 2448×2048 pixels, which is resized to 512×512 pixels when training and evaluating our model.

The deep models are trained with the Adam optimizer, whose exponential decay rates of the 1st and 2nd order moment estimates are 0.9 and 0.999, respectively. The training epoch and batch size are 300 and 6, respectively. The learning rate is initialized as 0.001. Data augmentation is beneficial for the generalization ability. Before each optimization step, the random augmentation is applied, including hue change, brightness change, saturation change, contrast change, left-right flip, and up-down flip. The hardware configuration includes Intel Xeon Silver 4214R CPU and NVIDIA RTX3090 GPU.

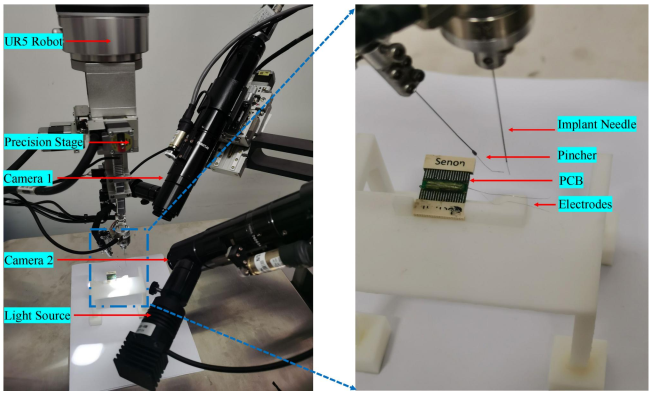

Fig. 5. Implantation robot system.

The Percentage of Correct Keypoint (PCK) metric is used to evaluate the localization results. PCK reports the percentage of correct localization results that fall within a given distance around the ground-truth keypoint coordinates, namely,

$$\text{PCK}=\frac{1}{M}\sum_{\text{i}=1}^{M} 1(d_i < t)$$

where M is the number of test samples; d_i is the distance between the predicted keypoints and the ground-truth of sample i. t=20pixel is a distant threshold. Besides, the Mean Pixel Error (MPE) is used as another evaluation metric. MPE computes the mean distance error between the predicted keypoint and the ground-truth keypoint, namely,

$$\text{MPE} = \frac{1}{M}\sum_{\text{i}=1}^{M} d_i$$

Similarly, Percentage of Correct Angle (PCA) and Mean Degree Error (MDE) are calculated in terms of direction angle to evaluate the angle estimation performance, as given by,

$$\text{PCA}=\frac{1}{M}\sum_{\text{i}=1}^{M} 1(\alpha_i < \gamma)$$

$$\text{MDE} = \frac{1}{M}\sum_{\text{i}=1}^{M} \alpha_i$$

where α_i is the difference between predict angle and ground-truth angle of sample i. γ=3° is the degree threshold.

C. *Ablation Experiments*

The baseline approach uses Resnet-34 without CBAM as encoder of U-shaped network and uses standard 3×3 convolution layer instead of DCN module. DCN, CBAM and DCN&CBAM are added to the baseline for comparison to investigate their effectiveness. We run a series of ablation experiments and the evaluation results are shown in Table I. Note the percentage results are calculated with 100 test samples, so that the percentages have no non-zero decimals.

Compared to the baseline, DCN improves the PCK of Needle from 89.0% to 92.0% and reduces the MPE of Electrode from 20.99 pixel to 13.76 pixel. The traditional convolution kernels have fixed shape and cannot adapt to different situations, while DCN can change the sampling position involved in convolution and improve the shape adaptiveness. Compared to the baseline, CBAM improves the PCA of Electrode from 94.0% to 96.0% and reduces the MDE of Electrode from 1.77 pixel to 1.52 pixel. CBAM introduces attention mechanism in channel and spatial dimension, so that the deep neural network can learn the area that needs attention

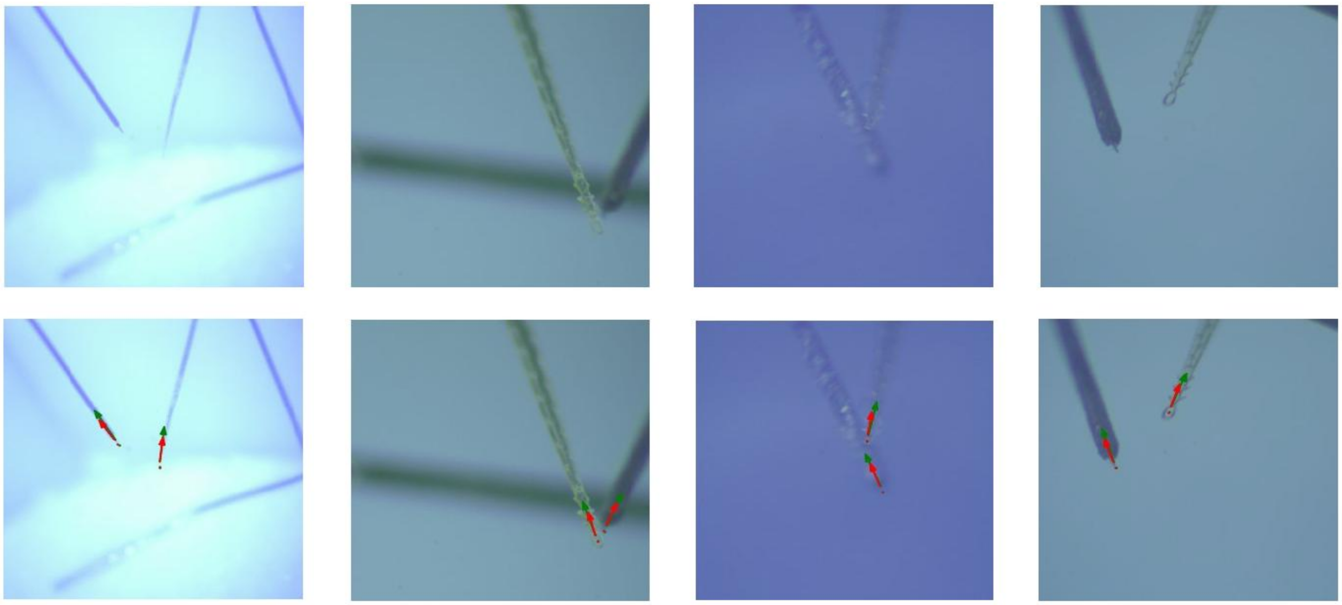

Fig. 5. Keypoint localization and angle estimation results of our model. The green points and arrows represent ground truths of keypoint and direction. Red points and arrows are the model's prediction.

TABLE I. ABLATION EXPERIMENTS RESULTS

Architecture	PCK (%)		PCA (%)		MPE (pixel)		MDE (degree)	
	Electrode	Needle	Electrode	Needle	Electrode	Needle	Electrode	Needle
Baseline	90.0	89.0	94.0	94.0	20.99	14.24	1.77	1.76
with CBAM	91.0	92.0	96.0	95.0	15.19	13.85	1.52	1.72
with DCN	91.0	91.0	95.0	95.0	13.76	13.92	1.54	1.62
with DCN & CBAM	**94.0**	**93.0**	**98.0**	**97.0**	**11.12**	**13.58**	**1.52**	**1.44**

TABLE II. COMPARISON EXPERIMENTS RESULTS

Model	PCK (%)		PCA (%)		MPE (pixel)		MDE (degree)		Time (ms)
	Electrode	Needle	Electrode	Needle	Electrode	Needle	Electrode	Needle	
G-RMI [12]	87.0	87.0	84.0	90.0	12.77	14.14	2.44	2.07	37.8
Hourglass [13]	92.0	91.0	76.0	92.0	11.22	14.16	2.64	1.96	15.6
SRNet [14]	90.0	89.0	91.0	92.0	12.21	14.07	1.98	1.89	**9.2**
Mask RCNN [15]	91.0	92.0	92.0	94.0	11.54	13.74	1.87	1.76	46.1
Ours	**94.0**	**93.0**	**98.0**	**97.0**	**11.12**	**13.58**	**1.52**	**1.44**	28.6

in each new image, focus on important features and suppress unnecessary features. The proposed model with both DCN and CBAM performs the best under PCK, PCA, MPE and MDE metrics.

D. *Comparison Experiments*

We compare the proposed KLAE-Net with four relevant models including G-RMI [12], Hourglass [13], SR-Net [14] and Mask RCNN [15]. The comparison experiments results are shown in Table II. Overall, the proposed KLAE-Net presents better performance on keypoint localization dataset than all compared methods. Besides, the inference speed of KLAE-Net satisfies the real-time requirements. The keypoint localization and angle estimation results of our model are visualized in Fig. 5. Although the imaging condition, object pose and lens magnification are varying, KLAE-Net can obtain the keypoints and directions correctly in most cases. When magnification changes, 99.0% keypoints and directions can be obtained correctly. When object defocus, 97.0% keypoints and directions can be obtained correctly. When partial occlusion occurs, 98.0% keypoints and directions can be obtained correctly. When illumination changes, 99.0% keypoints and directions can be obtained correctly.

E. *Alignment Control Experiment*

The trajectory of implant needle tip relative to electrode ring center in image space during the alignment control is shown in Fig. 6. It can be seen that under the guidance from microscopic vision, the implant needle is moved by the manipulator and gradually approaches the electrode ring center in image space, within ~10 steps. When the keypoints' position error is approximately zero, the alignment error in 3-D Cartesian space is also reduced to near zero. Note that when the image position error is under 20pixel, manual fine-tuning is conducted to finish the final high-precision alignment. The whole needle-electrode alignment process is

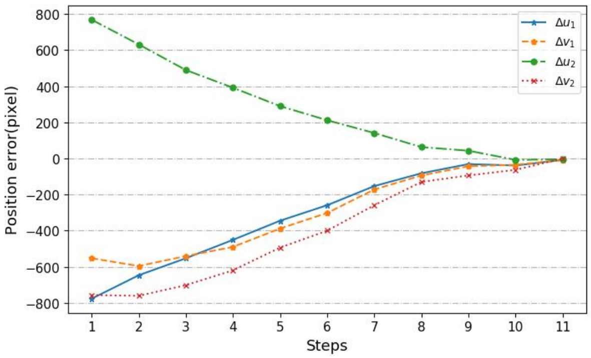

Fig. 6. Position error trajectory during needle-electrode alignment control.

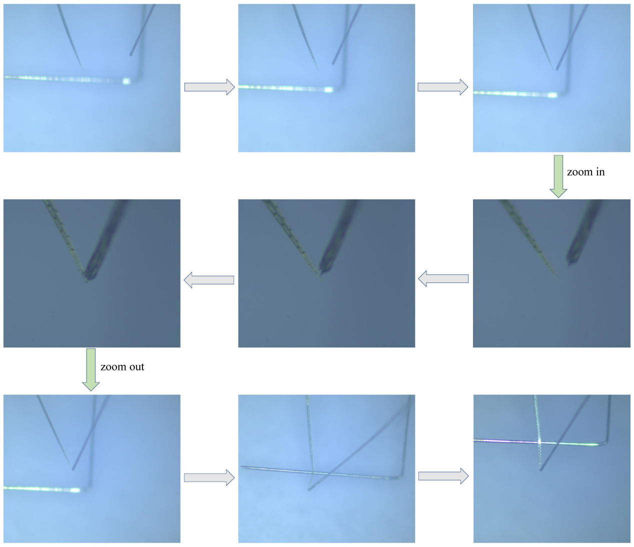

Fig. 7. Needle-electrode alignment process.

shown in Fig. 7. Firstly, the implant needle coarsely approaches the electrode end. Then the two microscopic cameras both switch to high magnification, and the alignment control is executed. After the needle and electrode is aligned, the needle tip moved along its axis to hook the flexible electrode.

VI. CONCLUSION

In this paper, the KLAE-Net is proposed to realize precise visual localization of micro flexible electrode and tiny implant needle for robotic electrode implantation. KLAE-Net is capable of predicting the coordinates and inclined angle of electrode and needle in real time. By training with numerous images under various conditions, the trained KLAE-Net has good generalization ability under the magnification, occlusion, defocus, and illumination condition changes. To realize the efficient alignment control of implant needle and flexible electrode, the keypoints predicted by KLAE-Net are used to represent the image position error and the image Jacobian matrices are used to map the image error to the position error in Cartesian space. According to the visual errors, the alignment controller outputs motion order to the manipulator so that the needle tip moves towards the electrode ring until the alignment is finished. In the future, we will improve the robustness and precision of keypoint localization, so that the alignment cab be executed fully automatically and reliably.

REFERENCES

[1] S. G. Mason and G. E. Birch, "A general framework for brain-computer interface design", *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, pp. 70-85, Mar. 2003.

[2] G. Buzsáki, C. A. Anastassiou and C. Koch, "The origin of extracellular fields and currents—EEG ECoG LFP and spikes", *Nature Rev. Neurosci.*, vol. 13, no. 6, pp. 407-420, 2012.

[3] V. S. Polikov, P. A. Tresco and W. M. Reichert, "Response of brain tissue to chronically implanted neural electrodes", *J. Neurosci. Methods*, vol. 148, no. 1, pp. 1-18, Oct. 2005.

[4] S. P. Lacour, S. Benmerah, E. Tarte, et al., " Flexible and stretchable micro-electrodes for in vitro and in vivo neural interfaces ", *Med. Biol. Eng. Comput.*, vol. 48, no. 10, pp. 945-954, 2010.

[5] C. Xie, J. Liu, T. M. Fu, et al., "Three-dimensional macroporous nanoelectronic networks as minimally invasive brain probes", *Nature Mater.*, vol. 14, no. 12, pp. 1286-1292, Dec. 2015.

[6] T. D. Kozai and D. R. Kipke, "Insertion shuttle with carboxyl terminated self-assembled monolayer coatings for implanting flexible polymer neural probes in the brain", *J. Neurosci. Meth.*, vol. 184, no. 2, pp. 199-205, 2009.

[7] Z. Xiang, S. C. Yen, N. Xue, et al., "Ultra-thin flexible polyimide neural probe embedded in a dissolvable maltose-coated microneedle", J. Micromech. Microeng., vol. 24, no. 6, pp. 065015, 2014.

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, et al., "Active shape models-their training and application", *Comput. Vis. Image Understanding*, pp. 38-59, Jan. 1995.

[9] G.J. Edwards, T.F. Cootes, and C.J. Taylor. "Face recognition using active appearance models", Eur. Conf. on Computer Vision (ECCV), pp. 581-595, 1998.

[10] P. Dollár, P. Welinder and P. Perona, "Cascaded pose regression", *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1078-1085, 2010.

[11] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[12] G. Papandreou, T. Zhu, N. Kanazawa, et al., "Towards accurate multi-person pose estimation in the wild", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4903-4911, Jul. 2017.

[13] A. Newell, K. Yang and J. Deng, "Stacked hourglass networks for human pose estimation", *Proc. Eur. Conf. Comput. Vis.*, pp. 483-499, 2016.

[14] T. Kurmann, P. M. Neila, X. Du, et al., "Simultaneous recognition and pose estimation of instruments in minimally invasive surgery", *Proc. Comput. Vis. Pattern Recognit.*, pp. 505-513, 2017.

[15] K. He, G. Gkioxari, P. Dollár, et al., "Mask R-CNN", *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 2980-2988, Oct. 2017.

[16] K. He, X. Zhang, S. Ren, et al., "Deep residual learning for image recognition", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770-778, Jun. 2016.

[17] S. Woo, J. Park, J.-Y. Lee, et al., "CBAM: Convolutional block attention module", *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 3-19, Sep. 2018.

[18] O. Ronneberger, P. Fischer and T Brox, "U-net: Convolutional networks for biomedical image segmentation", Proc. Medical Image Comput. Comp.-Assis. Interv. – MICCAI, vol. 9351, pp. 234-241, 2015.

[19] J. Dai, H. Qi, Y. Xiong, et al., "Deformable convolutional networks", *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 764-773, Oct. 2017.

[20] E. Musk and Neuralink, "An integrated brain-machine interface platform with thousands of channels", J. Med. Internet Res., vol. 21, no. 10, pp. e16194, 2019.

[21] L.-C. Chen, G. Papandreou, F. Schroff and H. Adam, "Rethinking atrous convolution for semantic image segmentation", arXiv:1706.05587, 2017, [online] Available: https://arxiv.org/abs/1706.05587.

[22] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing" in The Confluence of Vision and Control, New York:Springer-Verlag, vol. 237, pp. 66-78, 1998.