Preliminary Studies of a Model for a Robot that Creates an Interactive Communication with Elderly People to Satisfy Their Clothing Item Requests

Olivia Nocentini[1(✉)], Jaeseok Kim[1], and Filippo Cavallo[1,2]

[1] Department of Excellence in Robotics and AI, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, 56127 Pisa, Italy
olivia.nocentini@santannapisa.it

[2] Department of Industrial Engineering, University of Florence, Via Santa Marta 3, 50139 Firenze, Italy

Abstract. Robotic technology has become more and more sophisticated in the last epoch and a consequence of this revolution is the huge diffusion of robots. Currently, robots are no more relegated only to the industrial field, indeed they can be employed in houses, as assistants for several tasks, or in hospitals. In this paper, in particular, we are going to describe a possible implementation of an innovative model for an assistive robot that creates an interactive communication with elderly people to satisfy their clothing item requests. In detail, assistive robots need several skills during this kind of task, such as cloth classification, manipulation of clothes, and human-robot interaction. Concerning manipulation, the use of machine learning is increasing in this field, which develops fast and powerful grasping models and this is the reason why we would like to use one of them in the grasping phase. The reasons described before brought us to propose an innovative model for an assistive robot that satisfies clothing item requests of elderly people.

Keywords: Assistive robot · Manipulation of clothes · Convolutional neural network

1 Introduction

The interaction and collaboration between robots and humans are becoming more and more important in our society because of the increasing aging of the population. People prefer to live in their own homes for as long as possible, instead of being institutionalized in sheltered homes, or nursery homes when problems related to aging appear [2]. This is the reason why there is a growing necessity for new technologies that can assist the elderly in their daily living. Assistive robots can be used with elderly people in different scenarios: for prevention of early degeneration of cognitive abilities, the management of chronic

A. Bettelli et al. (Eds.): ForItAAL 2020, LNEE 884, pp. 73–84, 2022.
https://doi.org/10.1007/978-3-031-08838-4_5

diseases, fall prevention, maintaining social contacts, and the management of daily activities [6,8].

The reasons described above brought us to present this proposal whose aim is to implement an innovative model for an assistive robot that satisfies clothing item requests of elderly people. In particular, this model could be applied to a robot that helps elderly people in their own houses or in retirement homes.

In details, the main contributions of the proposal are the following:

- Creating a new large dataset of clothes useful for elderly people.
- Proposing from scratch a model for an assistive robot that does clothes classification using augmented features, grasps clothes using a specific neural network, and that can collect users' feedback.
- Testing the model using a real robot.

In detail, the paper is organized as follows: in Sect. 2, the state of the art is presented; in Sect. 3 methods and tools and in Sect. 4 discussions regarding the article are shown. In Sect. 5, the conclusions of the article are summarized.

2 State of the Art

The Assistive Robotics field is an area of growing interest [23] in which robots are used as tools to better assist human users with special needs [3], allowing them to perform some Activities of Daily Living (ADLs) by themselves. An interesting area of Assistive Robotics is the Robot-Assisted Dressing branch. This area is becoming extremely challenging because of the complex tasks, together with the difficulty to control the contact-rich interactions between a garment and the body part to be dressed [21]. Finally, other issues that have been not yet solved regarding the dataset of clothes that contains few items, and the communication with elderly people to satisfy their clothing item requests.

2.1 Clothes Classification

One of the fields in Robot-Assisted Dressing is the classification of clothes and it is a sub-field that remains still to be developed from many points of view, comparing to the field of grasping clothes. In [19], the work of the authors deals with the classification of garment categories including pants, shorts, shirts, T-shirts, and towels. The knowledge of the garment category is crucial for its robotic manipulation and their work focuses particularly on garments being held in a hanging state by a robotic arm and the input of their method is a set of depth maps taken from different viewpoints around the garment which are fused into a single 3D point cloud. In [5], a Baxter robot, a two-armed robot with an animated face, was used to dress a jacket onto a mannequin and human participants considering several combinations of user pose and clothing type (base layers), while recording dynamic data from the robot, a load cell, and an Inertial Measurement Unit (IMU). The authors also expand the analysis to include classification techniques such as decision trees and support vector machines using k-fold cross-validation. In [22], the authors presented a system for automatically extracting and classifying items in a pile of laundry. In detail, using only visual sensors, the robot identifies and extracts items sequentially from the pile. Finally, in [18] the authors proposed an attention-driven technique for tackling visual fashion clothes analysis in images, aiming to achieve clothing category classification and attribute prediction by producing regularised landmark layouts.

2.2 Grasping of Clothes

In this section, the selection of grasping points by the robot is shown. In [25], the authors described grasp points selection on an item of clothing randomly placed on a table, and in [14], the authors proposed a novel vision-based grasp points detection algorithm that can reliably detect the corners of a piece of cloth, using only geometric cues that are robust to variation in texture. In literature, several works also used Convolutional Neural Networks (CNNs) to train their dataset to find grasping points of clothes. In [17], the authors created a new dataset to train a CNN to improve the dressing assistance behaviour of a Pepper, a humanoid robot able to recognize faces and basic human emotions. In [16], the authors described a CNN that predicts the quality and pose of grasps at every pixel. This one-to-one mapping from a depth image overcomes the limitations of current deep-learning grasping techniques by avoiding discrete sampling of grasp candidates and long computation times. Finally, in [4] the problem of identifying optimal grasping poses for cloth-like deformable objects is addressed by means of a four-steps algorithm performing the processing of the data coming from a 3D camera.

3 Methods and Tools

In the following section, methods and tools for the implementation of the new model for an assistive robot, that helps elderly people in dressing tasks, are shown. The following block diagram describes the process before going into details (Fig. 1 and Fig. 2):

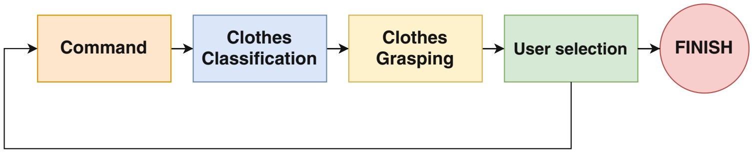


Fig. 1. Block diagram of the proposed model: it is composed of the command block followed by the clothes classification and grasping modules. After them, there is the user selection block that finishes the process which can be repeated.

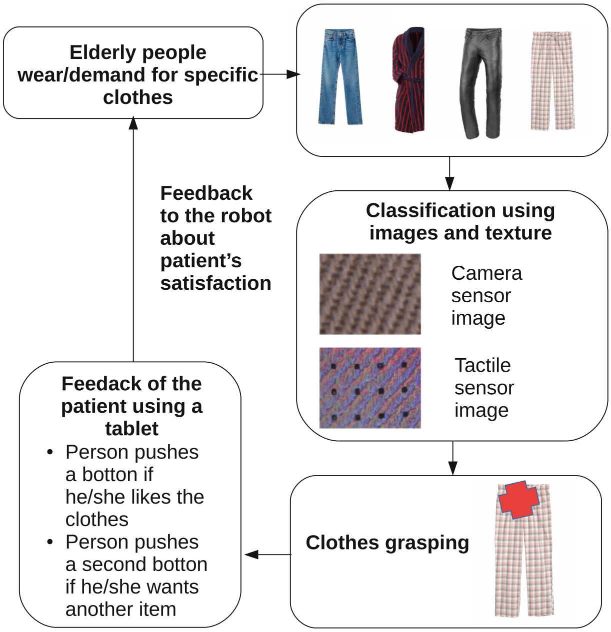


Fig. 2. Block diagram of the new network. In the first phase, the user demands for a specific cloth, then, clothes classification is done with or without augmented features. After classification, the robot grasps the cloth and gives it to the patient that gives a feedback based on what he/she receives.

3.1 Command

The command block represents what elderly people usually dress and one main goal of the model is to let the robot understanding the dressing needs of elderly people to dress them. To achieve this goal, first four types of clothes are put on a table: one pair of jeans, one leader trousers, one pajama, and one dressing gown of which the second two are usually worn by elderly people while the other two are worn by younger people. The novelty of the network consists of having a model of grasping clothes that has the goal to satisfy the dressing needs of elderly people.

3.2 Clothes Classification

This module is used to acquire images of clothes using an RGB-D camera. Both colours and depth of the camera are important in the grasping of clothes for having a better performance of the network. The images are acquired from the camera sensor that takes the RGB-D images of the clothes. The dataset is the split into 80% (training and dev set) and the remaining 20% part is the test set. The actual classification is done by concatenating the neural network to an MLP

(Multy-Layer Perceptron). In this module, the network is augmented concatenating some extra features to achieve a better clothes classification performance.

3.3 Clothes Grasping

In this module, the methodology of grasping clothes is chosen. We will use the Generative Grasping Convolutional Neural Network (GG-CNN) [16], modifying a little the internal part of it, to grasp clothes. This network predicts the quality and pose of grasps at every pixel. This one-to-one mapping from a depth image overcomes limitations of current deep-learning grasping techniques by avoiding discrete sampling of grasp candidates and long computation times. Additionally, GG-CNN is orders of magnitude smaller while detecting stable grasps with equivalent performance to current state-of-the-art techniques. The light-weight and single-pass generative nature of GG-CNN allows for closed-loop control at up 50 Hz, enabling accurate grasping in non-static environments where objects move and in the presence of robot control inaccuracies. The challenge of using this network is the fact that GG-CNN has problems with thin objects (like clothes), and so solving this issue could be a great contribution in the improvement of this network.

3.4 User Feedback

In this module, the elderly person takes the item that the robot has brought him and presses two buttons on the tablet that the robot has to answer the following questions using yes/no answers:

- Do you like the item that the robot brings you?
- Do you want another item?

Then according to which answers the elderly person gives on the tablet, the robot takes another cloth or stops its process.

The choice of using pressing buttons to answer specific questions is due to the low motor skills that elderly people have. In the tablet, the robot stores the information provided by elderly people to remember elderly people' satisfaction regarding the items that it gave them.

3.5 Hardware and Experimental Scenario

An important element is the choice of the robot: an idea is to test the grasping model using the assistive robot called Doro (Fig. 3) created by the Assistive Robotic Lab of Sant'Anna School of Advanced Studies [1]. This robot has an Xtion camera mounted on its head and has a Kinova 6-DoF arm with a hand as end-effector. In Fig. 4, a hypothetical experimental set-up scenario is shown.

Fig. 3. Doro robot assisting people in a house.

4 Preliminary Classification Methods

In this section, the preliminary results related to the classification method are presented. We used the Fashion-MNIST dataset as a preliminary dataset and we trained it on two different networks: a Convolutional Neural Network that we called FashionCNN, and a LeNet.

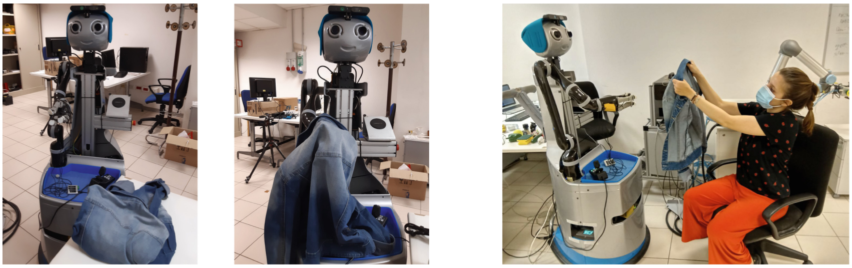
4.1 Proposed Networks Framework

For what concerns the clothes classification task, we took inspiration from [12] where the authors concatenated Histogram of Oriented Gradient (HOG) and color features to their network features to increase the performance of the network clothes classification. In our work, we decided to evaluate the performance of two different networks that use the Fashion-MNIST dataset with or without two features that are Canny and Sobel filters. We chose these two features since they can extract the shape of each cloth, providing an easier classification of the item.

FashionCNN. The first CNN used takes inspiration from the AlexNet network [10] and it has the following layers:

Two Sequential layers each consisting of the following layers:

- Convolution layer that has kernel size of 3 * 3, padding = 1 in 1st layer and padding = 0 in second one. Stride of 1 in both layers.
- Batch Normalization layer.
- Activation function
- Max Pooling layer with kernel size of 2 * 2 and stride 2.
- Flatten out the output for dense layer (a.k.a. fully connected layer).
- 3 Fully connected layers with different in/out features.
- 1 Dropout layer that has class probability p = 0.25.
  The input image is changing in a following way:
- First Convolution layer: input: 28 * 28 * 3, output: 28 * 28 * 32
- First Max Pooling layer: input: 28 * 28 * 32, output: 14 * 14 * 32



(a) Doro detects the item on the table (b) Doro manipulates the item (c) Doro gives the cloth to the person and the elderly person observes it

Fig. 4. Experimental set-up

– Second Conv layer: input: 14 * 14 * 32, output: 12 * 12 * 64
– Second Max Pooling layer: 12 * 12 * 64, output: 6 * 6 * 64
Final fully connected layer has 10 output features for 10 types of clothes.

LeNet. LeNet possesses the basic units of a CNN, such as convolutional layer, pooling layer and full connection layer. LeNet-5 consists of seven layers. The layer composition consists of 3 convolutional layers, 2 subsampling layers and 2 fully connected layers.

4.2 Features Extraction

The main goal of the experimental part is to obtain better clothes classification concatenating some features to the principal network features, as can be seen, in Fig. 5. In detail, we decided to concatenate to the network features, two other features that are the Canny edge detector and Sobel filter.

The augmented features that we used are the Canny edge detector and the Sobel Filter.

Canny Edge Detector. This is an operator that uses an algorithm to detect edges in images and it was developed by Canny in 1986. The Canny edge detection is a technique to extract useful structural information from different objects and reduce the amount of data to be processed. Among the edge detection methods developed so far, Canny edge detection algorithm is one of the most strictly defined methods that provide good and reliable detection.

Sobel Filter. The Sobel filter is used for edge detection and was developed by Sobel and Feldman. When we apply this mask to the image, it simply works like as first order derivate and calculates the difference of pixel intensities in an edge region.

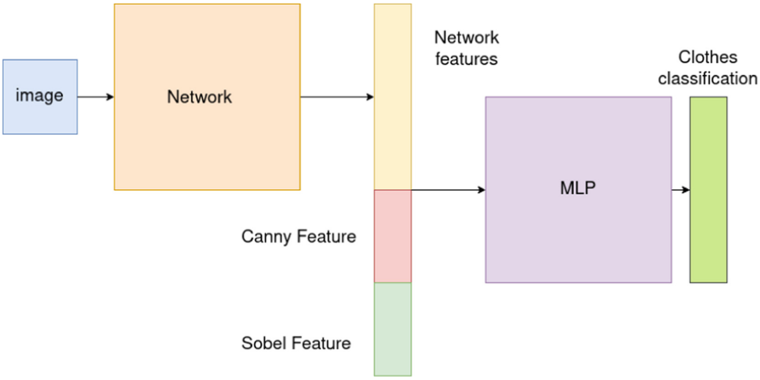


Fig. 5. Proposed model with the augmented features. The clothes image is processed by the neural network. Then, the features extracted from the network are concatenated to other features (Canny and Sobel) and fed into an MLP for classification.

4.3 Experiments

In the following paragraph, we compared the networks' performance with or without the Canny and Sobel features on the Fashion-MNIST Dataset.

The Fashion-MNIST Dataset. The Fashion-MNIST dataset is proposed as a more challenging replacement dataset for the MNIST dataset [11], it is based on the assortment on Zalando's website [24], and it is composed of pictures having a light-gray background. The original picture is resampled with multiple resolutions, e.g., large, medium, small, thumbnail, and tiny. The authors used the front look thumbnail images of 70,000 unique products to build Fashion-MNIST. In particular, white-color products are not included in the dataset as they have low contrast to the background. For the class labels, the authors used the silhouette code of the product.

Methods and Settings. In our work, we used the Adam optimizer together with the cross-entropy function as the loss function for training. The following equation describes the cross-entropy function:

$$Loss = -\sum_{s=1}^{n} y_i \log \hat{y}_i \tag{1}$$

where $\hat{y}_i$ is the i-th scalar value in the model output, y_i is the corresponding target value, and n is the number of scalar values in the model output. Experimentally, we set the learning rate as 0.01, the batch size as 128. We applied a 5-fold cross validation which divided the dataset into 80 % (training) and 20 % (test) set.

Results and Analysis. We first compared the loss and the accuracy of FashionCNN with and without Canny and Sobel features. From Table 1 can be seen that the average accuracy (86.8 %) without Canny and Sobel features is higher than the one with those filters (84.6 %). As a consequence the average loss (0.37) without features is less than the one with Canny and Sobel features (0.39).

Table 1. Losses and accuracies of FashionCNN with and without Canny and Sobel features

FashionCNN	F0	F1	F2	F3	F4	Mean
Loss without Canny and Sobel	0.37	0.35	0.35	0.44	0.36	**0.37**
Loss with Canny and Sobel	0.40	0.40	0.37	0.41	0.38	0.39
Accuracy without Canny and Sobel	87%	86%	88%	86%	87%	**86.8**%
Accuracy with Canny and Sobel	85%	84%	85%	86%	83%	84.6%

Secondly, we evaluated the loss and the accuracy of Lenet with and without Canny and Sobel features (Table 2). Concerning the accuracy average, we obtained in this case that the one with features is higher compared to the one without Canny and Sobel filters. (81.6% against 69.4 %). The losses were 0.51 and 0.81 with and without features respectively.

Table 2. Losses and accuracies of Lenet with and without Canny and Sobel features

Lenet	F0	F1	F2	F3	F4	Mean
Loss without Canny and Sobel	0.73	0.78	0.75	0.96	0.80	0.81
Loss with Canny and Sobel	0.52	0.50	0.51	0.50	0.51	**0.51**
Accuracy without Canny and Sobel	71%	69%	72%	67%	68%	69.4%
Accuracy with Canny and Sobel	81%	82%	81%	82%	82%	**81.6**%

From the results shown before, we achieved that the Canny and Sobel filters work and increase the performance of the Lenet, while there is no benefit using these features with the FashionCNN network. From our analysis, features from Lenet are not fully extracted so that sub-features, such as Canny and Sobel help to increase the performance. However, features from Fashion CNN are already sufficiently extracted. As a result, Canny and Sobel features are a factor of decreasing performance.

5 Future Work

In this paper, an innovative model for an assistive robot that satisfies clothing item requests of elderly people is presented. Although substantial advances in machine learning perception were made in the last decade, especially in object recognition [13] and action recognition [7], there's still a lack of systems that operate under diverse natural conditions and real-world time constraints that social interaction demands [26]. Moreover, most of the current social robots have been designed for interaction that lasts on the order of several minutes or hours but, on the contrary, human-human social interactions span months, and even years and this issue should be overcome in the future [26]. Going into details in the area of Assisted Dressing, even if this branch is increasing in importance, there are still some limitations in this area. Most of the studies on safety in robot-assisted dressing don't include tests with users and are limited to experiments on a mannequin. In [21], the authors used a styrofoam head with a dual-arm robot with attached anthropomorphic hand learning to put a knit cap. In [9], a mannequin was used to estimate Human-Cloth topological relationship using a depth sensor. Also in [20], the authors used a mannequin to test Reinforcement Learning for clothing assistance with a dual arm robot. Another issue for the future is that the majority of state of the art that works in robotic manipulation focuses on working with rigid objects, that either do not deform when they are grasped or have negligible deformation even if, deformable object manipulation has many important real-world applications (cloth folding, bed making, getting dressed) [15].

Finally, most of the papers on Robot-Assisted Dressing field do not take into consideration the clothing item requests and the feedback of people that they are dressing so this research project comes intending to solve the issues described before.

6 Conclusion

In this article, an innovative model for an assistive robot that satisfies clothing item requests of elderly people is presented.

Firstly, the state of the art is shown; then, methods and tools and preliminary classification methods with discussions elaborated from the state of the art are presented. This article came with the aim of solving issues presented in the discussions and tries to answer challenging problems of clothes manipulation in the Robot-Assisted Dressing field.

References

1. Bevilacqua, R., et al.: Robot-era project: preliminary results on the system usability. In: Marcus, A. (ed.) DUXU 2015. LNCS, vol. 9188, pp. 553–561. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20889-3_51
2. Broekens, J., Heerink, M., Rosendal, H., et al.: Assistive social robots in elderly care: a review. Gerontechnology **8**(2), 94–103 (2009)
3. Canal, G., Alenyà, G., Torras, C.: Adapting robot task planning to user preferences: an assistive shoe dressing example. Auton. Robots **43**, 1–14 (2018)

4. Caporali, A., Palli, G.: Pointcloud-based identification of optimal grasping poses for cloth-like deformable objects. In: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), vol. 1, pp. 581–586. IEEE (2020)
5. Chance, G., Jevtić, A., Caleb-Solly, P., Dogramadzi, S.: A quantitative analysis of dressing dynamics for robotic dressing assistance. Front. Robot. AI **4**, 13 (2017)
6. Flandorfer, P.: Population ageing and socially assistive robots for elderly persons: the importance of sociodemographic factors for user acceptance. Int. J. Popul. Res. **2012** (2012)
7. Herath, S., Harandi, M., Porikli, F.: Going deeper into action recognition: a survey. Image Vis. Comput. **60**, 4–21 (2017)
8. Johnson, D.O., et al.: Socially assistive robots: a comprehensive approach to extending independent living. Int. J. Soc. Robot. **6**(2), 195–211 (2014)
9. Koganti, N., Tamei, T., Matsubara, T., Shibata, T.: Estimation of human cloth topological relationship using depth sensor for robotic clothing assistance. In: Proceedings of Conference on Advances In Robotics, pp. 1–6. ACM (2013)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **25**, 1097–1105 (2012)
11. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
12. Li, R., Lu, W., Liang, H., Mao, Y., Wang, X.: Multiple features with extreme learning machines for clothing image recognition. IEEE Access **6**, 36283–36294 (2018). https://doi.org/10.1109/ACCESS.2018.2848966
13. Loncomilla, P., Ruiz-del Solar, J., Martínez, L.: Object recognition using local invariant features for robotic applications: a survey. Pattern Recogn. **60**, 499–514 (2016)
14. Maitin-Shepard, J., Cusumano-Towner, M., Lei, J., Abbeel, P.: Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 2308–2315. IEEE (2010)
15. Matas, J., James, S., Davison, A.J.: Sim-to-Real Reinforcement Learning for Deformable Object Manipulation. arXiv preprint arXiv:1806.07851 (2018)
16. Morrison, D., Corke, P., Leitner, J.: Closing the Loop for Robotic Grasping: A Real-Time, Generative Grasp Synthesis Approach. arXiv preprint arXiv:1804.05172 (2018)
17. Romeo, M., Cangelosi, A., Jones, R.: Developing a deep learning agent for HRI: dataset collection and training. In: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 1150–1155. IEEE (2018)
18. Shajini, M., Ramanan, A.: An improved landmark-driven and spatial-channel attentive convolutional neural network for fashion clothes classification. Vis. Comput. **37**, 1–10 (2020)
19. Stria, J., Hlavác, V.: Classification of hanging garments using learned features extracted from 3D point clouds. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5307–5312. IEEE (2018)
20. Tamei, T., Matsubara, T., Rai, A., Shibata, T.: Reinforcement learning of clothing assistance with a dual-arm robot. In: 2011 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids), pp. 733–738. IEEE (2011)
21. Twardon, L., Ritter, H.: Learning to put on a knit cap in a head-centric policy space. IEEE Robot. Autom. Lett. **3**(2), 764–771 (2018)

22. Willimon, B., Birchfield, S., Walker, I.: Classification of clothing using interactive perception. In: 2011 IEEE International Conference on Robotics and Automation, pp. 1862–1868. IEEE (2011)
23. Woiceshyn, L., Wang, Y., Nejat, G., Benhabib, B.: A socially assistive robot to help with getting dressed. In: 2017 Design of Medical Devices Conference, pp. V001T11A012–V001T11A012. American Society of Mechanical Engineers (2017)
24. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms (2017). arXiv preprint Cs.LG/1708.07747 32 (2017)
25. Yamazaki, K.: A method of grasp point selection from an item of clothing using hem element relations. Adv. Robot. **29**(1), 13–24 (2015)
26. Yang, G.Z., et al.: The grand challenges of science robotics. Sci. Robot. **3**(14), eaar7650 (2018)