

# HANDS18: Methods, Techniques and Applications for Hand Observation

Iason Oikonomidis<sup>3</sup>[0000-0002-9503-3723], Guillermo Garcia-Hernando<sup>1</sup>[0000-0003-3215-7857], Angela Yao<sup>5</sup>[0000-0001-7418-6141], Antonis Argyros<sup>2,3</sup>[0000-0001-8230-3192], Vincent Lepetit<sup>4</sup>[0000-0001-9985-4433], and Tae-Kyun Kim<sup>1</sup>[0000-0002-7587-6053]

<sup>1</sup> Imperial College London

<sup>2</sup> University of Crete

<sup>3</sup> Foundation for Research and Technology

<sup>4</sup> University of Bordeaux

<sup>5</sup> National University of Singapore

**Abstract.** This report outlines the proceedings of the Fourth International Workshop on Observing and Understanding Hands in Action (HANDS 2018). The fourth instantiation of this workshop attracted significant interest from both academia and the industry. The program of the workshop included regular papers that are published as the workshop’s proceedings, extended abstracts, invited posters, and invited talks. Topics of the submitted works and invited talks and posters included novel methods for hand pose estimation from RGB, depth, or skeletal data, datasets for special cases and real-world applications, and techniques for hand motion re-targeting and hand gesture recognition. The invited speakers are leaders in their respective areas of specialization, coming from both industry and academia. The main conclusions that can be drawn are the turn of the community towards RGB data and the maturation of some methods and techniques, which in turn has led to increasing interest for real-world applications.

**Keywords:** Hand Detection · Hand Pose Estimation · Hand Tracking · Gesture Recognition · Hand-Object Interaction · Hand Pose Dataset

## 1 Introduction

The Fourth International Workshop on Observing and Understanding Hands in Action was held in conjunction with the European Conference on Computer Vision 2018 (ECCV’18) on the 9th of September, 2018. It was held in the main building of the Technical University of Munich (TUM) in Arcisstraße 21, and specifically in the lecture hall Theresianum 606. The program of the workshop included five invited talks, six full papers, three extended abstracts, six invited posters, and a short award ceremony. This report presents in detail the proceedings of the workshop, and was not peer-reviewed.

---

The workshop website can be found at <https://sites.google.com/view/hands2018>



**Fig. 1.** Left: attending the talk of Professor Tamim Asfour. Right: awards photo. From left to right: Guillermo Garcia-Hernando, Angela Yao, Pavlo Molchanov, Umar Iqbal, Jan van Gemert, and Iason Oikonomidis.

## 2 Invited Talks

There were five invited talks in the workshop, the slides of which can be found on the workshop website.

- Christian Theobalt (Max Planck Institute for Informatics) presented an overview of his recent works on hand pose estimation. He presented several different scenarios: varying input modalities including depth and regular RGB, egocentric viewpoints, and hand-object interactions [28,26,22,21].
- Tamim Asfour (Karlsruhe Institute of Technology) spoke about the connection between stable grasps and humanoid locomotion states. After presenting an overview of his work on humanoid robotics, he presented the KIT whole-body human motion database [20] that can be used to efficiently solve humanoid locomotion problems.
- Andrew Fitzgibbon (Microsoft) talked about various optimizations over his line of work to achieve accurate real-time hand tracking performance. Among them, he focused on the Iterative Closest Point (ICP) algorithm and how it can be improved by, counter-intuitively, formulating a much larger optimization problem that includes both the model parameters and the correspondences in the same level of optimization [30]. The key insight is that each optimization iteration of the larger problem can take a bigger and more accurate step towards the optimum compared to standard ICP.
- Robert Wang (Facebook Reality Labs) talked about the process of acquiring ground truth data for hand pose estimation with markerless motion capture. After a discussion on the limitations of current approaches, he gave an overview of recent work at Facebook Reality Labs that aims to extract very accurate ground truth annotations using off-the-shelf equipment [13].
- Andrea Tagliasacchi (Google) gave an overview of his recent hand pose estimation works, including an approach for real-time hand tracking [29], a method to better model the shape of the hand [31], and a method to quickly and robustly personalized the hand model to the observed user [32].

### 3 Presented Works

There were three types of contributions accepted for presentation in the workshop. Specifically, there were regular contributions in the form of full papers (Accepted Papers, AP), extended abstracts (EA), and invited posters (IP). Regular papers were accepted based on a peer review process. Extended abstracts were evaluated and accepted by the organizers, and had a limit of three pages. Finally, works from the main ECCV'18 conference related to the aims and scope of the workshop were invited to be presented in the poster session of the workshop.

#### 3.1 Accepted Papers

The regular program of the workshop invited high quality original papers on the relevant areas. In total, there were seven submissions that were peer-reviewed by seventeen invited reviewers. Each paper was assigned three reviewers, aiming for at least two reviews per work. Through this review process, six of the seven submitted papers were accepted for publication in the workshop proceedings. In order of submission, the six Accepted Papers (AP) are:

**AP1: Hand-tremor Frequency Estimation in Videos** [25]. This paper deals with the problem of estimating the frequency of hand tremors in patients suffering from sensorimotor disorders such as Parkinson's disease. The authors used the highly successful 2D human keypoint estimation Pose Machine method by Wei et al. [35] to estimate 2D wrist positions over a sequence of frames. Using these positions, two alternative approaches are proposed for the estimation of the tremor frequency. The first was named the Lagrangian approach, in which a smooth trajectory was estimated from the sequence of 2D locations. Deviations of the hand from this smooth trajectory was then used to estimate the tremor frequency. For the second Eulerian method, again the same smoothed trajectory was used, but for this approach new image features are computed around the trajectory. An analysis of these new features yields the final frequency estimation. The two proposed methods were assessed on a new collected hand tremor dataset, TIM-Tremor, containing both static and dynamic tasks. The dataset contains data from 55 tremor patient recordings including accelerometer measurements that serve as ground truth, RGB images, and aligned depth data.

**AP2: DrawInAir: A Lightweight Gestural Interface Based on Fingertip Regression** [10]. "DrawInAir" proposes an in-air gestural recognition framework for Augmented Reality applications. The aim of this work is to enable real-time gesture recognition on lightweight devices such as a smartphones or other computationally constrained devices. The two main components of the framework are a fingertip localization module and a classification module that uses as input fingertip detection on subsequent frames and detects gestures. The first module is built using a fully convolutional network that outputs a heatmap of the fingertip. In contrast to common practice, an extra layer is used after

the heatmap generation, applying a differentiable spatial-to-numerical transform (DSNT) [24] to convert the heatmap to numerical coordinates using a soft-argmax operation. For gesture classification, a Bi-LSTM [12] approach is adopted; experimental evaluation shows that this performs better than standard LSTMs [15]. To experimentally evaluate the proposed method, the authors collect a new dataset called “EgoGestAR”.

**AP3: Adapting Egocentric Visual Hand Pose Estimation Towards a Robot-Controlled Exoskeleton** [4]. This paper also deals with patients suffering from motor impairments – here, the patient is assumed to have lost most of their motor abilities, and use an exoskeleton as a robotic grasp assistant. The aim of the system is to autonomously help the patient by estimating the hand pose and acting appropriately. Given that hand keypoint estimation methods usually assume that the hand is mostly free, with observed occlusions occurring from interaction with handled objects, the target scenario needs special treatment since the hand wears an exoskeleton. Towards this end, the authors propose a synthetic dataset that takes this fact into account, modeling the device and rendering hand poses with it. They adopt and adapt the approach of Wei et al. [35] comparing networks that are trained on data with and without the modeled device.

**AP4: Estimating 2D Multi-Hand Poses From Single Depth Images** [9]. This paper treats the problem of 2D hand keypoint detection of two hands in a single depth image. The authors use the Mask R-CNN object detector [14] to detect and segment the hands in the input image. Since Mask R-CNN can be generalized to multiple human bodies, or multiple hands pose estimation, a direct approach would be to train this pipeline on the target keypoints. However, as the authors state in the manuscript, minimal domain knowledge for human pose estimation is exploited so Mask R-CNN does not adequately model joint relationships. Moreover, another recent work [6] points out that, using this strategy, key points might not be localized accurately in complex situations. To address this limitations, the authors propose a Pictorial Structure [1] model-based framework. The authors evaluate the resulting system in two datasets that are generated from the single-hand datasets Dexter1 [27] and NYU hand pose dataset [33] by concatenating randomly selected left and right hand images.

**AP5: Spatial-Temporal Attention Res-TCN for Skeleton-based Dynamic Hand Gesture Recognition** [16]. This paper presents the Spatial-Temporal Attention Residual Temporal Convolutional Network (STA-Res-TCN) is to recognize dynamic hand gestures using skeleton-based input. The framework consists of an end-to-end trainable network that exploits both spatial and temporal information on the input data and applies an attention mechanism on both input modalities. This results in a lightweight but accurate gesture recognition system and is evaluated on two publicly available datasets [8,7].

**AP6: Task-Oriented Hand Motion Retargeting for Dexterous Manipulation Imitation** [2]. This paper treats the problem of retargeting already captured motion of a human hand on another hand embodiment, such as a dexterous anthropomorphic robotic hand. The formulation follows a task-oriented approach, namely the successful grasp of the manipulated object and formulates an objective taking the task goal into account. They proceed to learn a policy network using generative adversarial imitation learning. Experiments show that this approach achieves a higher success rate on the grasping task compared to a baseline that only retargets the motion using inverse kinematics.

### 3.2 Extended abstracts

Apart from the regular papers, the workshop also had original contributions in the form of extended abstracts, with the goal of including high-potential but preliminary works. These works were presented as posters in the poster session but are not published as part of the workshop proceedings. Of the four submissions, three were chosen for acceptance by the program chairs. In order of submission, the three extended abstracts (EA) are:

**EA1: Model-Based Hand Pose Estimation for Generalized Hand Shape with Spatial Transformer Network.** Recent work [39] has proposed using a hand kinematics model as a layer of a deep learning architecture, with the goal of making integrating differentiable coordinate transformations to enable end-to-end training. A limitation of that approach is the fact that the kinematics model has fixed parameters, making the resulting network specific for a single hand and not generalizing well to other hands. The authors of this work [19] extend the previous approach by adapting the kinematics parameters to the observed hand. A Spatial Transformer Network is also applied to the input image, which is shown by the experimental evaluation to be beneficial.

**EA2: A New Dataset and Human Benchmark for Partially-Occluded Hand-Pose Recognition During Hand-Object Interactions from Monocular RGB Images.** This work [3] proposes a dataset for pose estimation of partially occluded hands when handling an object. The authors collect a dataset that consists of a variety of images of hands grasping objects in natural settings. A simple strategy enables the recording of both occluded and un-occluded images of the same grasps. The error of human annotation is evaluated using this dataset.

**EA3: 3D Hand Pose Estimation from Monocular RGB Images using Advanced Conditional GAN.** This work [23] presents a method to estimate the 3D position of hand keypoints using as input a monocular RGB image. The authors propose to decompose the problem in two stages: the first estimates a depth map from the input RGB image using a cycle-consistent GAN architecture

[40]. The second stage employs a network based on Dense nets [17] to regress the joint positions using as input the estimated depth map.

### 3.3 Invited Posters

The organizers invited the following works from the main ECCV’18 conference related in aim to the workshop to be presented in the poster session:

- IP1: “HandMap: Robust Hand Pose Estimation via Intermediate Dense Guidance Map Supervision” [36].
- IP2: “Point-to-Point Regression PointNet for 3D Hand Pose Estimation” [11].
- IP3: “Joint 3D tracking of a deformable object in interaction with a hand” [34].
- IP4: “Occlusion-aware Hand Pose Estimation Using Hierarchical Mixture Density Network” [37].
- IP5: “Hand Pose Estimation via Latent 2.5D Heatmap Regression” [18].
- IP6: “Weakly-supervised 3D Hand Pose Estimation from Monocular RGB Images” [5].

## 4 Awards

Two works were given awards sponsored by Facebook Reality Labs. The best paper award was decided by the program chairs, while the best poster award was decided in a vote from selected workshop attendants including organizers, invited speakers and topic experts.

The best paper award was given to the work “Hand-tremor Frequency Estimation in Videos” by Silvia L Pintea, Jian Zheng, Xilin Li, Paulina J.M. Bank, Jacobus J. van Hilten and Jan van Gemert. Apart from a solid technical contribution, the work shows the applicability of the methods and techniques related to the scope of this workshop in the aid of real-world problems. The best poster award was given to the work “Hand Pose Estimation via Latent 2.5D Heatmap Regression”, by Umar Iqbal, Pavlo Molchanov, Thomas Breuel, Juergen Gall and Jan Kautz. New works on hand pose estimation are increasingly turning again to regular RGB input. This work proposes a novel approach towards this end, and achieves state-of-the-art results.

## 5 Discussion

Table 1 provides an overview of the presented works: rows correspond to individual works (AP for Accepted Paper, EA for Extended Abstract, and IP for Invited Poster) and columns to work traits. Specifically, “RGB” is marked in works that use regular RGB images as input, “Depth” for ones that use depth data, and “Skeletal Data” for works that use as input an existing estimation of the keypoints of interest. The trait “Application” is marked for works that solve

**Table 1.** Overview of the presented works. AP stands for Accepted Paper, EA for Extended Abstract, and IP for Invited Poster. Works are listed in the order they were presented in the respective sections.

	RGB	Depth	Skeletal Data	Application	Dataset	Egocentric	Gesture
AP1			x	x	x		
AP2	x			x	x	x	
AP3				x		x	
AP4		x					
AP5			x				x
AP6			x	x			
EA1		x					
EA2					x		
EA3	x						
IP1		x					
IP2		x					
IP3		x		x	x		
IP4		x					
IP5	x						
IP6	x						
Total	4	6	3	5	4	2	1

real-world problems, “Dataset” for works that propose a new dataset, “Egocentric” for works that assume an egocentric observation of the hand, and “Gesture” for works that tackle the problem of recognizing hand gestures.

One conclusion that can be drawn is that the technical level of the related systems is reaching production-grade performance, with five of the fifteen works being applications. While depth-based works are still the norm, monocular RGB input is increasingly common. Furthermore, new datasets are being proposed for increasingly complex, real-world scenarios, and including stand-alone RGB input. Towards this end, relevant scenarios involve occlusions due to hand-object interactions, other environmental factors, haptics, and robotic learning methods such as imitation and reinforcement learning. All of these scenarios should be oriented towards applications that require high precision estimates. Some of these points were mentioned in the invited talks of the workshop, and current trends on datasets point in these directions.

Two main categories of hand pose estimation approaches have been identified in the relevant literature. Discriminative approaches, based on learning, directly map observations to output hand poses. On the other hand, generative approaches, often based on computer graphics techniques, synthesize observations that are then compared to inputs. Then, through optimization, the hand pose that most accurately matches the observation is identified. These two classes of approaches have had small overlap over the last years. There is still much to be done towards the integration of these approaches, despite the long line of research on both categories, and also towards their integration. The resulting, third category should combine the advantages of both categories in so

called hybrid approaches. Some of the invited talks (Andrew Fitzgibbon, Andrea Tagliasacchi) focused on generative approaches while others (Christian Theobalt, Robert Wang) focused more on discriminative or hybrid approaches. Towards this integration, Andrea Tagliasacchi suggested the use of hand segmentation in a generative approach as a potential approach.

Generative approaches are focusing on more efficient methods for adaptive hand models, efficient model representations, and optimization strategies. A potentially useful observation is the fact that all modern generative approaches formulate and optimize differentiable objective functions. It is conceivable then that some of them could be used directly as loss functions in training neural networks. On the learning front, approaches that use semi-supervised learning or weak supervision [5] can potentially play a big role in the immediate future. Also, techniques that enable end-to-end network training such as Spatial Transformer Networks [19], Differentiable Spatial to Numerical Transform modules [10,18] and kinematic layers [39] are evidently useful.

Developed methods and techniques are already being applied to solve real-world problems in healthcare, robotics, and AR/VR [25,10,2]. Other candidate application domains include: automotive environments, both regular and autonomous, for gesture-based interactions with the car. The surgery room for training, monitoring and aiding operations. Laboratory monitoring to record the interactions of hands and objects, and therefore, the experimental procedure. Overall, all aspects of human activity involving the manipulation of physical or virtual objects are potential candidates.

In connection to the integration of different approaches, challenging new datasets will prove useful towards the assessment of hybrid approaches. Furthermore, they can help highlight the strengths and weaknesses of discriminative and generative approaches. Following previous cases [38], a goal for the next editions of this workshop is to organize a challenge towards this end.

## References

1. Andriluka, M., Roth, S., Schiele, B.: Pictorial Structures Revisited: People Detection and Articulated Pose Estimation. *Proceedings of the IEEE International Conference on Computer Vision* (2017)
2. Antotsiou, D., Garcia-Hernando, G., Kim, T.K.: Task-Oriented Hand Motion Retargeting for Dexterous Manipulation Imitation. In: *Proceedings of the Fourth International Workshop on Observing and Understanding Hands in Action* (2018)
3. Barbu, A., Myanganbayar, B., Mata, C., Dekel, G., Ben-Yosef, G., Katz, B.: A new dataset and human benchmark for partially-occluded hand-pose recognition during hand-object interactions from monocular RGB images. In: *Extended Abstract Presentation at the Fourth International Workshop on Observing and Understanding Hands in Action* (2018)
4. Baulig, G., Gulde, T., Curio, C.: Adapting Egocentric Visual Hand Pose Estimation Towards a Robot-Controlled Exoskeleton. In: *Proceedings of the Fourth International Workshop on Observing and Understanding Hands in Action* (2018)



5. Cai, Y., Ge, L., Cai, J., Yuan, J.: Weakly-supervised 3D Hand Pose Estimation from Monocular RGB Images. In: European Conference on Computer Vision. pp. 1–17 (2018)
6. Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., Sun, J.: Cascaded Pyramid Network for Multi-Person Pose Estimation (2017). <https://doi.org/10.1109/CVPR.2018.00742>
7. De Smedt, Q., Wannous, H., Vandeborre, J.P., Guerry, J., Le Saux, B., Filliat, D.: SHREC'17 Track: 3D Hand Gesture Recognition Using a Depth and Skeletal Dataset (2017). <https://doi.org/10.2312/3dor.20171049>
8. De Smedt, Q., Wannous, H., Vandeborre, J.P.: Skeleton-Based Dynamic Hand Gesture Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops pp. 1206–1214 (2016). <https://doi.org/10.1109/CVPRW.2016.153>
9. Duan, L., Shen, M., Cui, S., Guo, Z., Oliver, D.: Estimating 2D Multi-Hand Poses From Single Depth Images. In: Proceedings of the Fourth International Workshop on Observing and Understanding Hands in Action (2018)
10. Garg, G., Hegde, S., Perla, R., Jain, V., Vig, L., Hebbalaguppe, R.: DrawInAir: A Lightweight Gestural Interface Based on Fingertip Regression. In: Proceedings of the Fourth International Workshop on Observing and Understanding Hands in Action (2018)
11. Ge, L., Ren, Z., Yuan, J.: Point-to-Point Regression PointNet for 3D Hand Pose Estimation. In: European Conference on Computer Vision (2018)
12. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM networks. Proceedings of the International Joint Conference on Neural Networks (5), 2047–2052 (2005). <https://doi.org/10.1109/IJCNN.2005.1556215>
13. Han, S., Liu, B., Wang, R., Ye, Y., Twigg, C.D., Kin, K.: Online optical marker-based hand tracking with deep labels. ACM Transactions on Graphics **37**(4), 1–10 (2018). <https://doi.org/10.1145/3197517.3201399>
14. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Iccv 2017 (2017)
15. Hochreiter, S., Unger Schmidhuber, J.: Long Short-Term Memory. Neural Computation (8), 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
16. Hou, J., Wang, G., Chen, X., Xue, J.H., Zhu, R., Yang, H.: Spatial-Temporal Attention Res-TCN for Skeleton-based Dynamic Hand Gesture Recognition. In: Proceedings of the Fourth International Workshop on Observing and Understanding Hands in Action (2018)
17. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 pp. 2261–2269 (2017). <https://doi.org/10.1109/CVPR.2017.243>
18. Iqbal, U., Molchanov, P., Breuel, T., Gall, J., Kautz, J.: Hand Pose Estimation via Latent 2.5D Heatmap Regression. In: European Conference on Computer Vision (2018)
19. Li, S., Wöhlke, J., Lee, D.: Model-based Hand Pose Estimation for Generalized Hand Shape with Spatial Transformer Network. In: Extended Abstract Presentation at the Fourth International Workshop on Observing and Understanding Hands in Action (2018)
20. Mandery, C., Terlemez, Ö., Do, M., Vahrenkamp, N., Asfour, T.: The KIT whole-body human motion database. In: International Conference on Advanced Robotics, ICAR, pp. 329–336 (2015). <https://doi.org/10.1109/ICAR.2015.7251476>

21. Mueller, F., Bernard, F., Sotnychenko, O., Mehta, D., Sridhar, S., Casas, D., Theobalt, C.: GANerated Hands for Real-time 3D Hand Tracking from Monocular RGB. In: CVPR 2018 (2018)
22. Mueller, F., Mehta, D., Sotnychenko, O., Sridhar, S., Casas, D., Theobalt, C.: Real-time Hand Tracking under Occlusion from an Egocentric RGB-D Sensor. arXiv preprint arXiv:1704.02201 (2017). <https://doi.org/10.1109/ICCV.2017.131>
23. Nguyen, L.H., Quan, L.M., Kim, Y.G.: 3D Hand Pose Estimation from Monocular RGB Images using Advanced Conditional GAN. In: Extended Abstract Presentation at the Fourth International Workshop on Observing and Understanding Hands in Action (2018)
24. Nibali, A., He, Z., Morgan, S., Prendergast, L.: Numerical Coordinate Regression with Convolutional Neural Networks (2018)
25. Pintea, S.L., Zheng, J., Li, X., Bank, P.J.M., van Hilten, J.J., van Gemert, J.: Hand-tremor Frequency Estimation in Videos. In: Proceedings of the Fourth International Workshop on Observing and Understanding Hands in Action (2018)
26. Sridhar, S., Mueller, F., Zollhöfer, M., Casas, D., Oulasvirta, A., Theobalt, C.: Real-time joint tracking of a hand manipulating an object from RGB-D input. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 9906 LNCS, pp. 294–310 (2016). [https://doi.org/10.1007/978-3-319-46475-6\\_19](https://doi.org/10.1007/978-3-319-46475-6_19)
27. Sridhar, S., Oulasvirta, A., Theobalt, C.: Interactive markerless articulated hand motion tracking using RGB and depth data. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2456–2463 (2013). <https://doi.org/10.1109/ICCV.2013.305>
28. Sridhar, S., Rhodin, H., Seidel, H.P., Oulasvirta, A., Theobalt, C.: Real-time hand tracking using a sum of anisotropic gaussians model. In: Proceedings - 2014 International Conference on 3D Vision, 3DV 2014. pp. 319–326 (2015). <https://doi.org/10.1109/3DV.2014.37>
29. Tagliasacchi, A., Schröder, M., Tkach, A., Bouaziz, S., Botsch, M., Pauly, M.: Robust Articulated-ICP for Real-Time Hand Tracking. In: Computer Graphics Forum (2015)
30. Taylor, J., Luff, B., Topalian, A., Wood, E., Khamis, S., Kohli, P., Izadi, S., Banks, R., Fitzgibbon, A., Shotton, J., Bordeaux, L., Cashman, T., Corish, B., Keskin, C., Sharp, T., Soto, E., Sweeney, D., Valentin, J.: Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences. *ACM Transactions on Graphics* **35**(4), 1–12 (2016). <https://doi.org/10.1145/2897824.2925965>
31. Tkach, A., Pauly, M., Tagliasacchi, A.: Sphere-meshes for real-time hand modeling and tracking. *ACM Transactions on Graphics* **35**(6), 1–11 (2016). <https://doi.org/10.1145/2980179.2980226>
32. Tkach, A., Tagliasacchi, A., Remelli, E., Pauly, M., Fitzgibbon, A.: Online generative model personalization for hand tracking. *ACM Transactions on Graphics* **36**(6), 1–11 (2017). <https://doi.org/10.1145/3130800.3130830>
33. Tompson, J., Stein, M., Lecun, Y., Perlin, K.: Real-Time Continuous Pose Recovery of Human Hands Using Convolutional Networks. *ACM Transactions on Graphics (SIGGRAPH 2014)* (5), 1–10 (2014). <https://doi.org/10.1145/2629500>
34. Tsoli, A., Argyros, A.A.: Joint 3D Tracking of a Deformable Object in Interaction with a Hand. In: European Conference on Computer Vision (2018)
35. Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: Proceedings of the IEEE Computer Society Conference

- on Computer Vision and Pattern Recognition. pp. 4724–4732 (2016).  
<https://doi.org/10.1109/CVPR.2016.511>
36. Wu, X., Finnegan, D., Neill, O., Yang, Y.L.: HandMap: Robust Hand Pose Estimation via Intermediate Dense Guidance Map Supervision. In: European Conference on Computer Vision (2018)
  37. Ye, Q., Kim, T.K.: Occlusion-aware Hand Pose Estimation Using Hierarchical Mixture Density Network. In: European Conference on Computer Vision (2018)
  38. Yuan, S., Garcia-Hernando, G., Stenger, B., Moon, G., Chang, J.Y., Lee, K.M., Molchanov, P., Kautz, J., Honari, S., Ge, L., Yuan, J., Chen, X., Wang, G., Yang, F., Akiyama, K., Wu, Y., Wan, Q., Madadi, M., Escalera, S., Li, S., Lee, D., Oikonomidis, I., Argyros, A., Kim, T.K.: Depth-Based 3D Hand Pose Estimation: From Current Achievements to Future Goals. In: CVPR 2018 (2018)
  39. Zhou, X., Wan, Q., Wei, Z., Xue, X., Wei, Y.: Model-based deep hand pose estimation. IJCAI International Joint Conference on Artificial Intelligence pp. 2421–2427 (2016)
  40. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. Proceedings of the IEEE International Conference on Computer Vision pp. 2242–2251 (2017).  
<https://doi.org/10.1109/ICCV.2017.244>