



Exploiting and Guiding User Interaction in Interactive Machine Teaching

Zhongyi Zhou

Interactive Intelligent Systems Lab., The University of Tokyo
Tokyo, Japan
zhongyi@iis-lab.org

ABSTRACT

Humans are talented with the ability to perform diverse interactions in the teaching process. However, when humans want to teach AI, existing interactive systems only allow humans to perform repetitive labeling, causing an unsatisfactory teaching experience. My Ph.D. research studies Interactive Machine Teaching (IMT), an emerging field of HCI research that aims to enhance humans' teaching experience in the AI creation process. My research builds IMT systems that exploit and guide user interaction and shows that such in-depth integration of human interaction can benefit both AI models and user experience.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; • **Computing methodologies** → *Computer vision; Machine learning*.

KEYWORDS

interactive machine teaching, saliency map, deictic gestures, in-situ annotation, dataset, data diversity

ACM Reference Format:

Zhongyi Zhou. 2022. Exploiting and Guiding User Interaction in Interactive Machine Teaching. In *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22 Adjunct)*, October 29–November 2, 2022, Bend, OR, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3526114.3558529>

1 INTRODUCTION

Artificial Intelligence (AI) is changing the world by assisting various applications that benefit humans' life [10, 29, 30]. To create new AI applications, researchers usually teach AI new concepts [33] (e.g., what a cat looks like) by training Machine Learning (ML) models [20, 25] on large-scale datasets labeled by human annotators [4]. Despite the importance of data labeling, I argue that it is only one of the massive interactions humans are talented at performing during the teaching process. Humans' teaching interaction provide rich information about the concept that humans want to teach, and thus such interaction should be exploited and encouraged when humans interact with AI [24]. However, due

to the lack of interactive technologies that support diverse teaching behaviors, existing systems mainly constrained human teachers in performing repetitive labeling, causing an unsatisfactory teaching experience.

Interactive Machine Teaching (IMT) [16, 19] is an emerging field of research in HCI that aims to enhance humans' teaching experience during the creation of Machine Learning (ML) models. Different from developing models through data collection and programming that require professional expertise, typical Vision-based IMT (V-IMT) systems allow users to teach a model by demonstrations, which are intuitive behaviors people perform in teaching. For example, Teachable Machine [2] allows users to teach vision-based ML classifiers by demonstrating the objects of different classes in front of a camera. After the teaching process, the system automates all machine learning processes, and the user can further assess the created model, as well as decide whether they need to perform another iteration of teaching-and-assessing processes [5]. Based on such standard system designs, recent studies show that IMT systems should further engage non-experts by providing guidance [11, 21] as well as exploiting more human interactions beyond labeling [16, 24] during the teaching process. For example, Fiebrink *et al.* [7] found that users wanted more information in the assessment interface so that they could understand “*where and how the model was likely to make mistakes*”. Hong *et al.* [11] further highlighted the importance of guidance in supporting users to decide “*what to show in the teaching set*”.

My Ph.D. research focuses on exploiting and guiding users' teaching interaction in V-IMT systems. My first project provided teaching guidance by enhancing users' interpretation of the trained models. I created an assessment interface with saliency map visualization that explains what portions of the images the model weighs heavily in the prediction. I further found that the model created by standard V-IMT systems may easily misinterpret a concept by highlighting unrelated features. My second project summarized the cause of the issue as a lack of fine-grained annotations of objects of interest that users want to teach. To address this issue, I created a V-IMT system, called LookHere, that integrates object annotations into the teaching process by exploiting users' deictic gestures towards objects of interest. The user study shows that the in-situ object annotation achieved by exploiting humans' gestural interaction can significantly accelerate the teaching process without a noticeable model accuracy drop. In addition to exploiting user interaction, in my third (on-going) project, I propose a V-IMT that guides users to perform informative teaching. The teaching interface will visualize how different a given view can be from the existing teaching set in real time, and thus encourage users to cover a wide range of views in the teaching set.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

UIST '22 Adjunct, October 29–November 2, 2022, Bend, OR, USA

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9321-8/22/10.

<https://doi.org/10.1145/3526114.3558529>

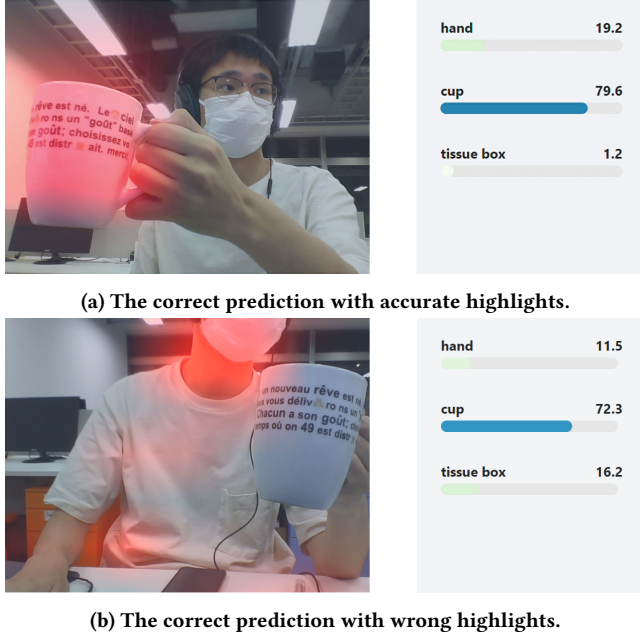


Figure 1: Example views of the model testing interface [31]. Bar visualization of confidence scores is shown on the right side. A camera view with heat-map overlays that explains the model prediction is shown on the left side. (a) shows an example with a correct explanation while the explanation in (b) is wrong.

2 MODEL ASSESSMENT VIA SALIENCY MAPS

The model assessment interface is a critical component of the IMT system. Using the interface, the user can be informed of the model performance, which decides whether and how they will teach the model to correct errors. Existing designs for model assessment mainly present confidence scores of each class as assistant data when the user interacts with the camera. Despite its usefulness, I argue that there is a lack of support for the user to interpret why the model makes each prediction. Intuitively, when a human teacher wants to assess whether a human student has fully understood a concept, the teacher may not only care about the student’s final answer but also whether the student can answer with appropriate reasons. In this work, I enhanced the model assessment interfaces in V-IMT through real-time saliency map visualizations.

Figure 1 illustrates an example view of my model assessment interface. In addition to the confidence scores shown on the right side, the interface also presents saliency maps in real time on the left side. The saliency maps help the user interpret why the model predicts a target category (the default is the one with the highest confidence score, i.e., “cup” in Figure 1a&1b). By examining the highlighted regions, the user can validate whether the ML model has correctly understood each concept by visualizing the correct regions. For example, in Figure 1a, the model provides a correct explanation of what “cup” is, whereas the model in Figure 1b mistakenly highlights the masked human face as the “cup”, showing the learning failure. Interestingly, both models show similarly high confidence ($\sim 70\%$) for the prediction, implying that the user may trust the model if there is no assistance from my saliency maps.

Such a case when the model is highly confident of a correct prediction but uses a totally wrong reference is not rare. This reveals a potential risk in prior V-IMT systems: the model created from the simplified teaching experience may not precisely learn a target concept. The main reason is that when the user demonstrates the object in front of a camera, the camera view may inevitably include other visual data than the target object. Therefore, what the user defines as, for instance, “cup”, is the full scenes of the images instead of portions of the images that represent “cup”. This reveals a teaching ambiguity issue in existing V-IMT systems, and future work should investigate how to support the user better clarify the object to teach.

3 GESTURE-AWARE TEACHING WITH IN-SITU ANNOTATIONS

One naïve approach to address the teaching ambiguity issue is to let the user annotate the object of interest in each teaching sample, emphasizing which object the model should learn. Although existing research on interactive annotation provided simplified interaction for the annotation process (e.g., by clicking [15, 22], sketching [17, 28], or mouse dragging [3], such post-hoc annotations inevitably bring extra workload that may diminish human teachers’ user experiences.

Can V-IMT systems integrate object annotations into teaching so that the user can express in-situ annotations through simple interaction? In this project, I exploited humans’ deictic gestures for achieving such integration and built a gesture-aware V-IMT system, called LookHere. The main intuition behind the system design is that humans can naturally interact with the object using deictic gestures in the teaching process. For example, during the teaching process, the user may hold the object or point at the object. Such gestures explicitly provide important cues on where the object of interest is. My system, LookHere, can intelligently capture such cues embedded with humans deictic gestures and achieves in-situ annotations in the teaching process.

3.1 LookHere

The key characteristic of LookHere is that it integrates the annotation process into human teaching by leveraging the user’s deictic gestures. Figure 2 shows an example view of the teaching interface in LookHere. Other than standard functions in other V-IMT systems [2, 8] (e.g., visualizing counts of teaching samples in each category), LookHere includes a function called *object highlights* to inform the user what regions of the camera view the system considers as the object to learn. The object highlights work in real time at approximately 28.3 fps using one GTX 2080Ti GPU. After the user clicks on the camera icon on the bottom-left corner of Figure 2, LookHere saves an image-label pair, as well as a segmentation mask which is visualized as object highlights on the teaching interface.

3.2 Gesture-aware Object Highlights

LookHere achieves the in-situ annotation using our gesture-aware object segmentation algorithm. To support real uses in V-IMT, the segmentation algorithm ought to be object-agnostic (i.e., without the constraints of what object can be segmented) because V-IMT systems allow users to teach a wide range of daily objects. LookHere

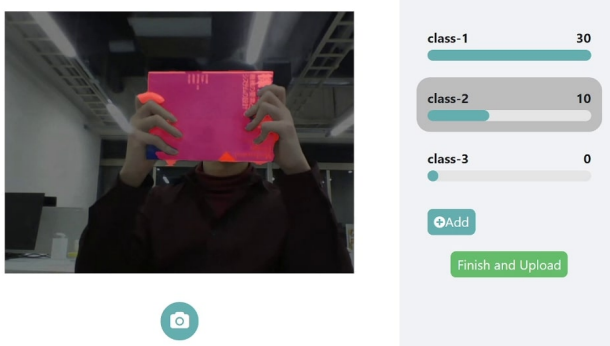


Figure 2: LookHere [32]. The camera view with object highlights is presented on the left side of the visualization of the count of teaching samples in each category.

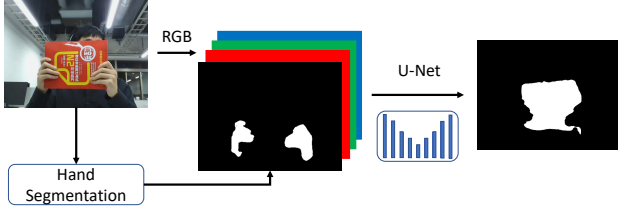


Figure 3: The workflow of the gesture-aware object-agnostic object segmentation algorithm [32]. The object highlights are created by using a U-Net and inputting the original RGB image attached to its hand segmentation result.

addresses this challenge through a gesture-aware algorithm as well as our customized dataset. Figure 3 summarizes the workflow of our algorithm. LookHere first performs a hand segmentation with a given RGB image. It can then feed both the RGB image and the hand segmentation mask into U-Net, which predicts a segmentation mask of the object specified by deictic gestures. The main intuition of algorithm design is that the hand segmentation should provide informative cues on where the object specified by the users’ gestures is, making it possible to perform object-agnostic segmentation.

3.3 HuTics

The most related dataset to the use case of LookHere is TEgO [13], and thus I tried training the algorithm on TEgO with augmented labels. The results show that the trained model is not robust enough to accurately various daily objects. Therefore, in this work, I collected a customized dataset, called HuTics, which included 2040 images collected from 170 people using human deictic gestures to interact with diverse daily-life objects. To cover a wide range of deictic gestures that humans may perform, I refer to Sauppe et al.’s taxonomy [18] and divide deictic gestures into four categories (i.e., exhibiting, pointing, presenting and touching). Each participant was required to use these four kinds of deictic gestures to interact with daily-life objects and took 12 photos in total (i.e., three photos for each kind of deictic gesture). I highly encouraged them to interact with a wide range of daily-life objects in their 12 photos. I then recruited five annotators to label the segmentation of the object.

We further used the data from 80% of the participants in HuTics (i.e., 1632 images from 136 people) for training and 20%

Table 1: The mean values and standard deviations of the task completion time and accuracies (classification and segmentation) across the four interface conditions.

		LookHere	NaïveIMT	Click	Contour
Time [s]		104 (44)	67 (10)	1,197 (228)	1,483 (407)
Acc.	Cls.	0.824 (0.158)	0.847 (0.190)	0.880 (0.141)	0.833 (0.159)
	Seg.	0.605 (0.153)	0.139 (0.095)	0.716 (0.167)	0.732 (0.151)

for testing. We found that the accuracy of our gesture-aware algorithm (Figure 3) trained on HuTics is 0.718. The accuracy of the same algorithm trained on TEgO with augmentation is 0.368, showing a large accuracy drop. Please see our full paper for details of the algorithm choice [32]. This demonstrates that HuTics plays a critical role in powering my object-agnostic segmentation algorithm compared to directly using other related datasets with augmentation.

3.4 Evaluations

To evaluate LookHere, we conducted a user study with 12 participants. None of them has experience in studying or working in the fields related to AI or ML. The purpose of this study is to understand the benefits of LookHere by comparing it with three baseline interface designs:

- *NaïveIMT*: This represents the most common design in current V-IMT systems [2, 8], in which participants only perform object demonstration without annotation.
- *Contour*: In addition to the teaching process with the naïve IMT system, this condition would involve a manual annotation procedure, using a contour-based tool [1], in a post-hoc manner.
- *Click*: This condition replaces the contour-based tool in the “Contour” condition with a click-based annotation method [23] to represent a simplified annotation process.

Participants were required to teach a vision-based classification model under each given interface condition, and we measured three kinds of metrics in the study: 1) time consumption; 2) model accuracy; 3) NASA-TLX [9]. Note that the model accuracy includes both classification accuracy as well as segmentation accuracy, aiming to test whether the model can not only predict correctly but also explain the prediction accurately.

Table 1 summarizes the results of time consumption and two types of accuracies. The results reflect that LookHere can enable a fast model creation experience without significantly sacrificing the model accuracy. NASA-TLX results (see detailed data in the full paper [32]) further show LookHere causes a lower cognitive burden compared to two conditions that require post-hoc annotations. Therefore, our user study demonstrates that LookHere can achieve a good balance between accuracy and workload for V-IMT due to our system designs that integrate annotations into teaching.

4 TEACHING GUIDANCE BY VISUALIZING DATA DIVERSITY IN REAL TIME

LookHere solves the issue of clarifying *what* (regions in the camera view) the user wants to teach by exploiting gestural interaction. In



Figure 4: Example images with visualization in *HuTics* dataset. *HuTics* covers four kinds of deictic gestures to objects: exhibiting (top-left), pointing (top-right), presenting (bottom-left) and touching (bottom-right). The hands and objects of interest are highlighted in blue and green, respectively. Note that there is no human annotation of hand segmentation, and the blue regions are from the prediction of Li et al.’s method [14].

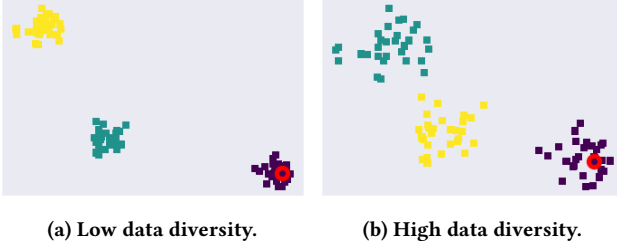


Figure 5: Example views of visualization in the teaching interface. (b) describes a distribution of teaching data that has higher data diversity than (a).

addition to exploiting human interaction, I argue that IMT research should also motivate human interaction by guiding users *how* to perform informative teaching. In practice, it is highly challenging to guide users’ teaching behaviors in real time because it usually takes time for a system to judge whether the data are valuable for training or not. To compute the data value [27], a system needs to train a back-end model on the data, which is a very time-consuming procedure, particularly for deep learning models. Fails and Olsen [6] also highlighted this issue when they first introduced interactive machine learning research. They further argued that researchers should use lightweight models like decision trees (DTs) as the back-end models instead of Neural Networks (NNs) so that users can receive rapid feedback on their teaching behaviors.

In this on-going project, I plan to create a system that provides real-time teaching guidance when the user is teaching a deep learning model. To achieve this goal, this study challenges a widely acknowledged design principle that the system provides feedback *after* the model finishes training on the full training set [2, 31]. Instead, I argue that the quality of teaching data can be computed *before* the time-consuming training process, which overcomes the main bottleneck of rapid feedback designs in IMT research. This work considers data diversity as a key factor that can boost the informativeness of teaching, which is a commonly acknowledged heuristic that can benefit ML models [12]. Compared to the data value, data diversity is much easier to compute. More importantly, the computation of data diversity is independent of the choice of back-end models, making it possible for the real-time feedback in IMT system that trains deep models at the back end.

Figure 5 illustrates two examples of how I plan to visualize the data diversity in the teaching interface when the user teaches a three-way classifier. Each rectangle in the visualization represents an image recorded in the system, and the color represents its classification label. The circle highlighted by the red edge is the image that the user is currently teaching, which moves in real time with the change of the camera view. By continuously interacting with the circle, the user is encouraged to build each cluster to be as large and sparse as possible (i.e., Figure 5b is better than Figure 5a). For example, both circles in Figure 5b&5a are not ideal teaching data to be added to the teaching set since both fail to expand the coverage of the purple categories. I envision that such real-time visualization can motivate the users to present diverse views of the target objects, benefiting the machine learning process.

5 DISCUSSION AND FUTURE WORK

User interaction beyond labeling during the teaching process require more in-depth studies to benefit both user experience and AI models. My research studies user interaction in IMT systems by exploiting deictic gestures and guiding users’ object demonstration processes. In reality, humans use many interactions in the teaching process, and therefore more user interactions than those covered by this paper should be exploited and guided to enhance human-AI collaboration. For example, future work can study how to exploit gaze and verbal interaction that also contain rich information on the concepts users want to teach. On the other hand, users still encounter many challenges in which they have no idea how to perform effective teaching [11, 26]. Future work should study how to guide user interaction in these challenging scenarios (e.g., how systems can support users to correct a model that misinterprets a concept other than simply labeling more data).

It is important to note that this paper assumes that there is only one human teacher in the human-AI interaction. However, in practice, multiple users may teach models collaboratively. With more users engaged in the teaching process, new types of teaching interactions may emerge, bringing new research questions that requires investigation. Future research should observe these interactions and further study how the interactions can be exploited and guided to enhance collaboration.

REFERENCES

- [1] Dimitrios Bounias, Ashish Singh, Spyridon Bakas, Sarthak Pati, Saima Rathore, Hamed Akbari, Michel Bilello, Benjamin A Greenberger, Joseph Lombardo, Rhea D Chitalia, et al. 2021. Interactive Machine Learning-Based Multi-Label Segmentation of Solid Tumors and Organs. *Applied Sciences* 11, 16 (2021), 7488.
- [2] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. 2020. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3334480.3382839>
- [3] Chia-Ming Chang, Chia-Hsien Lee, and Takeo Igarashi. 2021. Spatial Labeling: Leveraging Spatial Layout for Improving Label Quality in Non-Expert Image Annotation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 306, 12 pages. <https://doi.org/10.1145/3411764.3445165>
- [4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [5] John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. *ACM Trans. Interact. Intell. Syst.* 8, 2, Article 8 (jun 2018), 37 pages. <https://doi.org/10.1145/3185517>
- [6] Jerry Alan Fails and Dan R. Olsen. 2003. Interactive Machine Learning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces* (Miami, Florida, USA) (IUI '03). Association for Computing Machinery, New York, NY, USA, 39–45. <https://doi.org/10.1145/604045.604056>
- [7] Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. 2011. Human Model Evaluation in Interactive Supervised Learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 147–156. <https://doi.org/10.1145/1978942.1978965>
- [8] Jules François, Baptiste Caramiaux, and Téo Sanchez. 2021. Marcelle: Composing Interactive Machine Learning Workflows and Interfaces. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (UIST '21). Association for Computing Machinery, New York, NY, USA, 39–53. <https://doi.org/10.1145/3472749.3474734>
- [9] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- [10] Hirotaka Hayashi, Anran Xu, Zhongyi Zhou, and Koji Yatani. 2021. Vision-Based Scene Analysis toward Dangerous Cycling Behavior Detection Using Smartphones. Association for Computing Machinery, New York, NY, USA, 28–29. <https://doi.org/10.1145/3460418.3479300>
- [11] Jonggi Hong, Kyungjun Lee, June Xu, and Hernisa Kacorri. 2020. Crowdsourcing the Perception of Machine Teaching. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376428>
- [12] Woojin Kang, In-Taek Jung, DaeHo Lee, and Jin-Hyuk Hong. 2021. Styling Words: A Simple and Natural Way to Increase Variability in Training Data Collection for Gesture Recognition. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445457>
- [13] Kyungjun Lee and Hernisa Kacorri. 2019. Hands Holding Clues for Object Recognition in Teachable Machines. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300566>
- [14] Peike Li, Yunqiu Xu, Yunchao Wei, and Yi Yang. 2020. Self-Correction for Human Parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020), 1–1. <https://doi.org/10.1109/TPAMI.2020.3048039>
- [15] Eric N Mortensen and William A Barrett. 1998. Interactive segmentation with intelligent scissors. *Graphical models and image processing* 60, 5 (1998), 349–384.
- [16] Gonzalo A. Ramos, Christopher Meek, Patrice Y. Simard, Jina Suh, and Soroush Ghorashi. 2020. Interactive machine teaching: a human-centered approach to building machine-learned models. *Hum. Comput. Interact.* 35, 5–6 (2020), 413–451. <https://doi.org/10.1080/07370024.2020.1734931>
- [17] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. "GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Trans. Graph.* 23, 3 (aug 2004), 309–314. <https://doi.org/10.1145/1015706.1015720>
- [18] Allison Sauppé and Bilge Mutlu. 2014. Robot Deictics: How Gesture and Context Shape Referential Communication. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany) (HRI '14). Association for Computing Machinery, New York, NY, USA, 342–349. <https://doi.org/10.1145/2559636.2559657>
- [19] Patrice Y. Simard, Saleema Amershi, David M. Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, and John Wernsing. 2017. Machine Teaching: A New Paradigm for Building Machine Learning Systems. <https://doi.org/10.48550/ARXIV.1707.06742>
- [20] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. <https://doi.org/10.48550/ARXIV.1409.1556>
- [21] Alison Smith-Renner, Ron Fan, Melissa Birchfield, Tongshuang Wu, Jordan Boyd-Graber, Daniel S. Weld, and Leah Findlater. 2020. No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376624>
- [22] Konstantin Sofiiuk, Ilia Petrov, Olga Barinova, and Anton Konushin. 2020. f-brs: Rethinking backpropagating refinement for interactive segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8623–8632.
- [23] Konstantin Sofiiuk, Ilia A. Petrov, and Anton Konushin. 2021. Reviving Iterative Training with Mask Guidance for Interactive Segmentation. [arXiv:2102.06583 \[cs.CV\]](https://arxiv.org/abs/2102.06583)
- [24] Nicole Sultanum, Soroush Ghorashi, Christopher Meek, and Gonzalo Ramos. 2020. A Teaching Language for Building Object Detection Models. Association for Computing Machinery, New York, NY, USA, 1223–1234. <https://doi.org/10.1145/3357236.3395545>
- [25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- [26] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300509>
- [27] Jinsung Yoon, Serkan Arik, and Tomas Pfister. 2020. Data valuation using reinforcement learning. In *International Conference on Machine Learning*. PMLR, 10842–10851.
- [28] Jing Zhang, Xin Yu, Aixuan Li, Peipei Song, Bowen Liu, and Yuchao Dai. 2020. Weakly-Supervised Salient Object Detection via Scribble Annotations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [29] Zhihang Zhong, Mingdeng Cao, Xiao Sun, Zhirong Wu, Zhongyi Zhou, Yinqiang Zheng, Stephen Lin, and Imari Sato. 2022. Bringing Rolling Shutter Images Alive with Dual Reversed Distortion. <https://doi.org/10.48550/ARXIV.2203.06451>
- [30] Zhongyi Zhou, Anran Xu, and Koji Yatani. 2021. SyncUp: Vision-Based Practice Support for Synchronized Dancing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 143 (Sept. 2021), 25 pages. <https://doi.org/10.1145/3478120>
- [31] Zhongyi Zhou and Koji Yatani. 2021. Enhancing Model Assessment in Vision-Based Interactive Machine Teaching through Real-Time Saliency Map Visualization. In *The Adjunct Publication of the 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (UIST '21). Association for Computing Machinery, New York, NY, USA, 112–114. <https://doi.org/10.1145/3474349.3480194>
- [32] Zhongyi Zhou and Koji Yatani. 2022. Gesture-aware Interactive Machine Teaching with In-situ Object Annotations. <https://doi.org/10.48550/ARXIV.2208.01211>
- [33] Xiaojin Zhu. 2015. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29.