Overview of key technologies of human-computer interaction

Yuyang Cheng[1, a], Benjamin Fang[2, b], Kai Lu[3, c], Xingyu Wei[4, d], Yifan Wang[5, e], Michael Yang[6, f], Xingchen Zhou[7, g], Zhenrong Zhang[8, h]

[1] School of Oxford International College of Chengdu, Chengdu 610000, China;

[2] Weston Highschool, 444 Wellesley Street Weston, MA 02493

[3] University of electronic science and technology of China, Chengdu, 610000, China

[4] New Hampton School, New Hampton03256, USA

[5] University of science and technology of China, Anhui, 230000, China.

[6] Brewster Academy, 569 Quaking Aspen Lane Windsor, CA95492

[7] Zhejiang University, Haining 314400, China

[8] University of Manchester, Manchester, UK

[a]MAME2013@163.com, [b]Benjamin.fang7@gmail.com, [c]18067355266@163.com, [d]Weix024@newhampton.org, [e]wyf2@mail.ustc.edu.cn, [f]michael0330@sina.com, [g]xingchen.18@intl.zju.edu.cn, [h]zzr534889920@163.com

Abstract. In the 21st century, computers are developing faster and faster. Human-computer interaction(HCL) is defined as a multidisciplinary field of study which focuses on the interaction between humans and computer technology. The advancement of HCL makes a great contribution to nowadays' society as human beings gradually perceive themselves in a technological world that is in a time of high-speed development.The situation of human-computer interaction is becoming more and more common, and computers have more and more intelligent factors. The latest computer technology has been applied to various disciplines, which have developed rapidly and achieved remarkable research results. In this paper, we summarizes the core technologies involved in human-computer interaction (HCI), mainly including human-computer interaction psychology, flexible screen technology, target recognition and classification, and expounds the current situation and development trend of these technologies.

Keywords: Human-computer interaction, Cognitive psychology, Flexible screen, Object Detection and Classification.

1. An overview of cognitive psychology in human-computer interaction

Human-computer interaction (HCL) is defined as a multidisciplinary field of study which focuses on the interaction between humans and computer technology. The advancement of HCL makes a great contribution to nowadays' society as human beings gradually perceive themselves in a technological world that is in a time of high-speed development. Aside from that, there is growing concern about the correlation between HCL and cognitive psychology. Cognitive psychology, in the Cambridge dictionary, refers to a combination of neuropsychology, cognitive psychology such as brain function and learning theory within a rehabilitation structure.

Don Norman, a cognitive psychologist, computer scientist, as well as an author, published a book named The Design of Everyday Things in 2013. Then a Norman door is firstly supposed in this book. This is referred to as a door where the wrong signal is given and instruction should be offered to correct it. It can also be defined as a door that the actual utilization of it is opposite to people's original thought. In other words, confusing doors is usually called "Norman doors." For instance, some doors(Norman doors) in a building usually need a signal of 'push' or 'pull' to remind the users. Otherwise, users might have an uncomfortable experience by using it incorrectly. [1]

Several guidelines are supposed by Don Norman when designing something. They are Discoverability, Feedback, Affordances, constraints, signifiers, mappings and Conceptual Models respectively. So-called seven basic principles of design. To partially discuss some guidelines, we could know how these seven sections are combined. Firstly, discoverability refers to a situation where people immediately know the operating model of an invention after they looked at it. This can be applied to many kinds of computer systems instead of being restricted to one specific area. [2] For example, sometimes users might confuse that whether they should tap once or twice to transfer a document when there are no instructions in specific in our electronic devices. Whereas normally we do not know the feedback of something by understanding what happened and why it happened when talking about feedback. [1] Affordances should be the most crucial section as it displays the way of using an object. Therefore, we could know how an object is operated. [2] Inventions with these seven principles are commonly known as User/ Human-centered design and it acts as the basis of how engineers work today. It is also a circulation that involves observation, idea generation, prototyping and testing. [1] By observing the results of the multiple tests, we could continually make improvements until it works. According to a book named The Future of Design Education Initiative, Elissaveta M. Brandon concludes that it is more crucial for designers to focus attention on a radically comprehensive aspect instead of narrowly looking at the surface of inventions. [3] Generally, it is more necessary to invent an object with good accessibility and usability rather than its appearance.

Don Norman had been a vice president in the 1990s at Apple. At that time, he set a special term which is user experience. This plays an essential role which acts as a platform that connects both users and the designers and allows them to share each other's opinions. Additionally, user experiences have become popularized in today's society. For example, it is easy to see that almost all the stores whatever it is an online one or an offline one, all provide this kind of service. This enhances the influence of HCL due to more frequent communication between people and computer technology, the development of it can also improve the efficiency of the development of HCL.

2. Overview of flexible screen technology in HCI

2.1 Flexible screen requirements for HCI

Since Google developed the first smart glasses in 2013, people's interest in flexible smart devices has been heightened, followed by a series of inventions such as smart wristbands and curved TVS. The continuous development of flexible screen technology is inseparable from its specific needs.

First of all, the advent of wearable human-computer interaction devices has higher requirements on the flexibility of the screen, which is expected to be suitable for various realistic environments and have the strongest anti-distortion ability as possible. In addition, the increasing demand for specific smart devices, such as smart helmets and smart glasses while riding, makes hard screens inadequate and may lead to unexpected safety hazards. At the same time, flexible screen devices will bring better visual experience to users, such as in the field of games and movies. These have aroused people's enthusiasm and strengthened the demand for flexible screens in human-computer interaction.

2.2 Flexible screen technology introduction, characteristics

The flexible screen usually consists of a flexible substrate and a display device layer structure formed on one side of the flexible substrate. For example, the structure of a flexible OLED display screen can include a flexible substrate, a plurality of conductive layers formed on the flexible substrate, and an insulating layer between the conductive layers. Flexible OLED has a unique manufacturing process, with a light material and an ultra-thin shape that allows it to stick stably to plastic or metal surfaces. If the plastic as a material, the weight of the display is lighter, wear resistance and bending will be better, while the traditional glass material is not used, so that it can withstand higher bending. There are also other attractive flexible screen materials, based on fibers and fabrics, which have the advantages of comfort, water resistance and 3d deformation.

However, it should be noted that although flexible screen technology has some achievements, people still have not realized the real flexible screen technology, currently facing many technical problems. At present, the flexible screen is curved rather than flexible, and there is no material that can really be folded as a whole. Also, OLED has some disadvantages, such as fear of water and oxygen. These factors often encountered in the environment into the device will affect the service life of the device, packaging technology also need to improve.

2.3 Flexible technology development and some applications.

In the early stage of flexible Display, two Korean LCD manufacturers, LG Display and Samsung Electronics, tried to bring the Display into the flexible era. On October 7, 2013, LG Display announced the start of mass production of its first flexible OLED panel, and in November it unveiled a smartphone with a curved Display. Samsung followed with its own flexible smartphones, but they were expensive.Then there are Google's smart glasses, Apple's smartwatch, and so on. These new inventions have prompted businesses to pay attention to the flexible screen device market. As consumers' expectations and the market push forward, OLED has become the next generation of products with high hopes as the scale and market share have grown. Researchers in Canada have unveiled the world's first truly flexible smartphone, which can fold and unfold freely. The enhanced visual experience solves the contradiction between large hd screen and portability, and brings a new way of human-computer interaction.

3. Overview of Object Detection in HCI

Based on the convolutional network and regression model stated above, the task of classification and localization of single object can be successfully achieved. However, when it comes to detecting multi-object in the same image, the task could be too complex and time-consuming for simply ordinary CNN, for it requires to extract lots of bounding boxes that might be of different size or overlapping with each other. Therefore, new frameworks have been introduced.

3.1 R-CNN

Raised by Ross Girshick from UC Berkeley, R-CNN (Region+CNN) is a milestone in applying the conventional CNN method to the complex object detection problems. The innovation of R-CNN is introducing region proposal method to extract every potential bounding boxes ahead of traditional CNN to enhance the performance of complex multi-object detection.

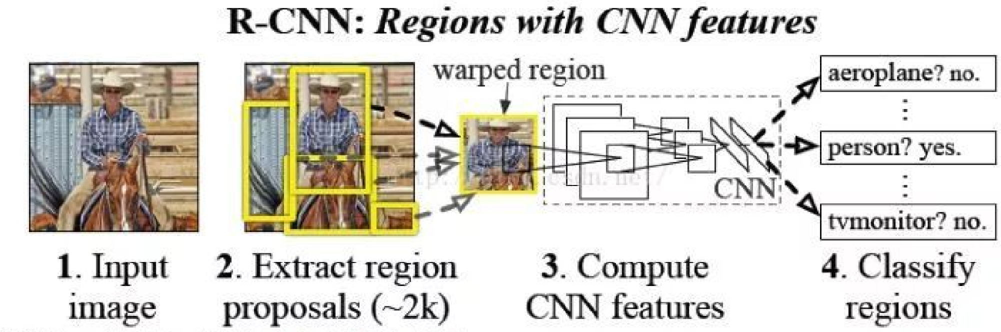

Figure 1. Graphic Illustration of R-CNN Architecture

The overall procedure can be illustrated as follow. First, use the Selective Search algorithm to extract nearly two thousand Region Proposals, which can also be called as Region of Interest (RoI), from the input image. Since the CNN model used in the original article contains a fully connected layer, this requires that the region proposal input to the network have the same size. Therefore, for each region, modify the size of the regions to a fixed scale and send them into a pre-trained and fine-tuned conventional CNN (like AlexNet, VGG) to extract their feature maps. Then these maps will be sent into and trained a SVM classifier, where can determine the category of the object in this candidate region. Each category corresponds to a SVM value. This stage will determine whether the potential object in region belongs to this specific category. If yes, then it is positive, otherwise it is negative. After SVM, we will get the class probability of these two thousand region proposals, then filter out undesired region proposals according to the category to get the surrounding as bounding box. And use the regression model to fine-tune the position of the candidate bounding box; for each category, train a linear regression model to determine whether the bounding box is perfect.

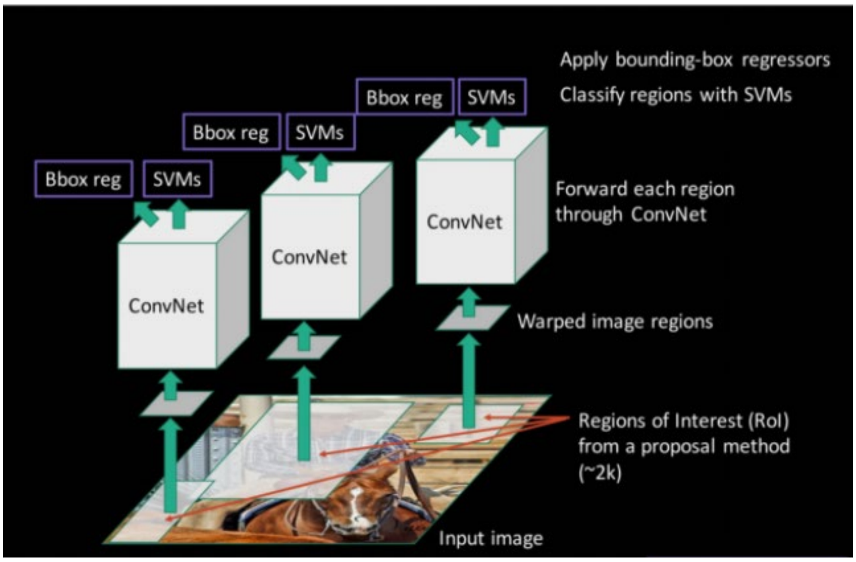

Figure 2. Region Proposal and SVMs Training

VOC 2010 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
DPM v5 [18]†	49.2	53.8	13.1	15.3	35.5	53.4	49.7	27.0	17.2	28.8	14.7	17.8	46.4	51.2	47.7	10.8	34.2	20.7	43.8	38.3	33.4
UVA [34]	56.2	42.4	15.3	12.6	21.8	49.3	36.8	46.1	12.9	32.1	30.0	36.5	43.5	52.9	32.9	15.3	41.1	31.8	47.0	44.8	35.1
Regionlets [36]	65.0	48.9	25.9	24.6	24.5	56.1	54.5	51.2	17.0	28.9	30.2	35.8	40.2	55.7	43.5	14.3	43.9	32.6	54.0	45.9	39.7
SegDPM [16]†	61.4	53.4	25.6	25.2	35.5	51.7	50.6	50.8	19.3	33.8	26.8	40.4	48.3	54.4	47.1	14.8	38.7	35.0	52.8	43.1	40.4
R-CNN	67.1	64.1	46.7	32.0	30.5	56.4	57.2	65.9	27.0	47.3	40.9	66.6	57.8	65.9	53.6	26.7	56.5	38.1	52.8	50.2	50.2
R-CNN BB	**71.8**	**65.8**	**53.0**	**36.8**	**35.9**	**59.7**	**60.0**	**69.9**	**27.9**	**50.6**	**41.4**	**70.0**	**62.0**	**69.0**	**58.1**	**29.5**	**59.4**	**39.3**	**61.2**	**52.4**	**53.7**

Figure 3. Detection average precision (%) on VOC 2010 test

3.2 Fast R-CNN

Raised in 2015, Fast R-CNN is an enhancement to previous model, proposed by Ross Girshick. Although successfully solved the issue of multi-object detection, R-CNN faced the problem: the training time is too long. The main reason to this issue is that R-CNN needs to train multiple times in different stages, and need to calculate the feature map separately through CNN for each region proposal, which leads to a longer overall time.

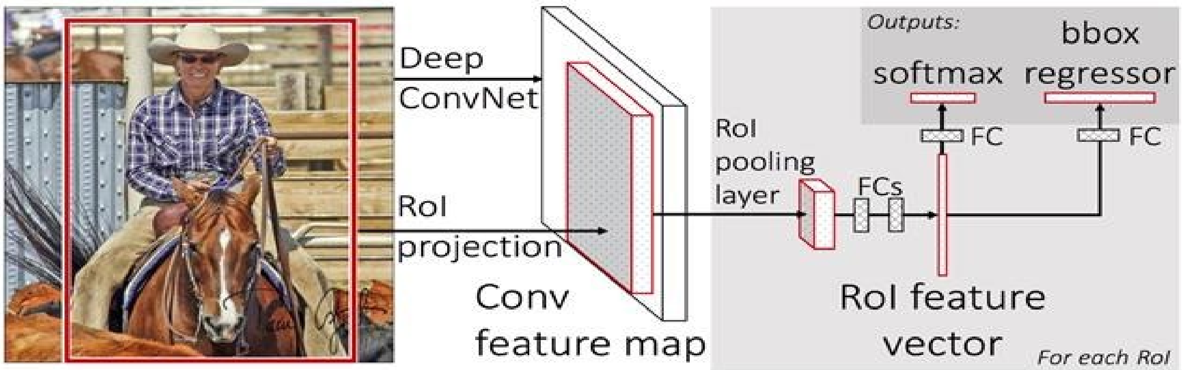

Figure 4. Graphic Illustration of Fast R-CNN Architecture

Therefore, Girshick used the idea from Spatial Pyramid Pooling (SSP Net), where after pooling, the input into the CNN can be of any scale compared to original fixed size. In SSP Net, convolution on the original image will be performed only once to obtain the feature map of the entire image; then find the mapping patch of each candidate region on the feature map, and input this patch as the final feature map of each candidate region into the subsequent layers. In the article, the author introduced the concept of ROI Pooling, which generally resembles SSP Net, and extract a fixed-dimensional feature representation for each region, and then perform category recognition through normal softmax. Also, in Fast-RCNN, the author cunningly put bbox regression into the CNN, to share convolution features with region classification.

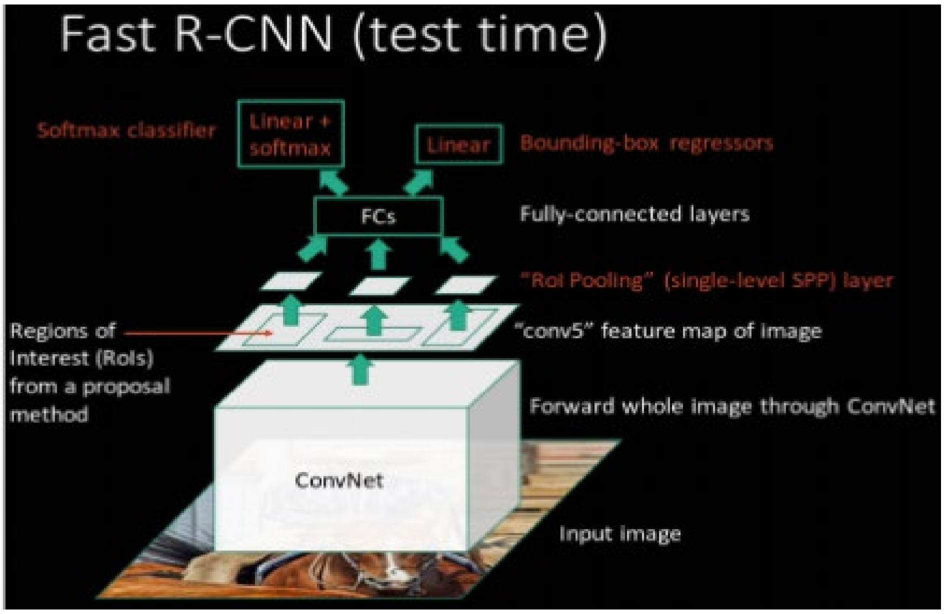

Figure 5. ROI Pooling and Shared Convolution Layer

Therefore, in Fast R-CNN, convolutional layers are shared, and not the candidate regions that entered into CNN as inputs, but the entire image. Therefore, the processing time can be greatly reduced.

	Fast R-CNN			R-CNN			SPPnet
	S	M	L	S	M	L	†L
train time (h)	**1.2**	2.0	9.5	22	28	84	25
train speedup	**18.3×**	14.0×	8.8×	1×	1×	1×	3.4×
test rate (s/im)	0.10	0.15	0.32	9.8	12.1	47.0	2.3
▷ with SVD	**0.06**	0.08	0.22	-	-	-	-
test speedup	98×	80×	146×	1×	1×	1×	20×
▷ with SVD	169×	150×	**213×**	-	-	-	-
VOC07 mAP	57.1	59.2	**66.9**	58.5	60.2	66.0	63.1
▷ with SVD	56.5	58.7	66.6	-	-	-	-

Figure 6. Runtime comparison between the same models in Fast R-CNN, R- CNN, and SPPnet.

3.3 Faster R-CNN

Compared to Fast R-CNN, Faster R-CNN continued to make progress to reduce the time cost. The author, Ross Girshick introduced Region Proposal Network(RPN), a CNN that can extract edges and find candidate region. In Faster R-CNN, the entire image is inputted into the first shared CNN to get the feature map like Fast R-CNN. After that, this feature map is sent to RPN to get the further feature of the candidate region before other procedure same as Fast R-CNN. By this, Faster R-CNN can save the time cost by selective search.

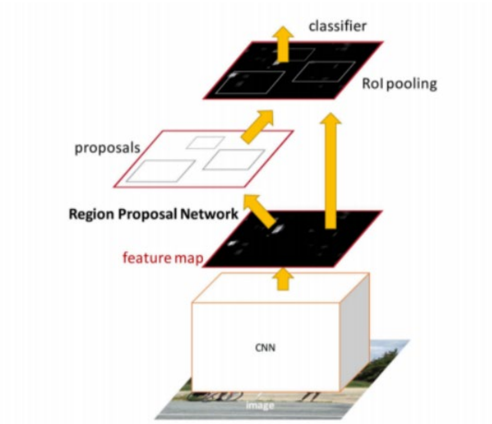

Figure 7. Faster R-CNN is a single, unified network for object detection

method	# box	data	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
SS	2000	12	65.7	80.3	74.7	66.9	46.9	37.7	73.9	68.6	87.7	41.7	71.1	51.1	86.0	77.8	79.8	69.8	32.1	65.5	63.8	76.4	61.7
SS	2000	07++12	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	**87.5**	80.5	80.8	72.0	35.1	68.3	**65.7**	80.4	64.2
RPN	300	12	67.0	82.3	76.4	71.0	48.4	45.2	72.1	72.3	87.3	42.2	73.7	50.0	86.8	78.7	78.4	77.4	34.5	70.1	57.1	77.1	58.9
RPN	300	07++12	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
RPN	300	COCO+07++12	**75.9**	**87.4**	**83.6**	**76.8**	**62.9**	**59.6**	**81.9**	**82.0**	**91.3**	**54.9**	**82.6**	**59.0**	**89.0**	**85.5**	**84.7**	**84.1**	**52.2**	**78.9**	65.5	**85.4**	**70.2**

Figure 8. Results on PASCAL VOC 2012 Test Set with Fast R-CNN Detectors and VGG-16

3.4 Classification tasks in human-computer interaction

Human-Computer Interaction (HCI) refers to the process of information exchange between human and computer using a certain dialogue language, in a certain way of interaction, to complete a certain task. HCI can be divided into several computer task in artificial intelligence field. For example machine learning (like time series) and deep learning task (like Nature Language Programming, Computer vision, Machine Vision). Machine vision is to use machines instead of human eyes to do measurement and judgment, attention is measurement and judgment.So in general, this is mainly used for factory inspection products and so on, mainly for some single scenarios.Computer vision refers to the use of cameras and computers instead of human eyes on the target identification tracking, etc.With the development of deep learning and neural network

in the image field,Deep learning plays an absolutely dominant role in computer vision. For how to help humans to complete the classification task in computer vision, before the development of neural network, computers need traditional image processing technologies such as Sobel and Lapacian to obtain feature points, and then put them into traditional machine learning such as SVM for classification. However, with the development of deep learning and format neural network, Deep learning has been able to better help people complete computer vision classification tasks in human-computer interaction. This section will introduce the development for HCI in computer vision classification tasks.

Image classification is a task to answer the question what the image is. This is no different from traditional machine learning classification tasksexcept that the input changes from numeric data to image data. Yann LeCun first proposed convolution - pooling - fully connected neural network structure in an article published in IEEE conference in 1998. Such 7-layer network structure was named LeNet5. The input image is classified after convolution - pooling - convolution - pooling - convolution - fully connected - fully connected. As a standard convolutional network structure, LeNet5 has exerted a great influence on later generations. In 2014, Google named the Inception structure GoogleNet to honor LeNet5’s contribution.

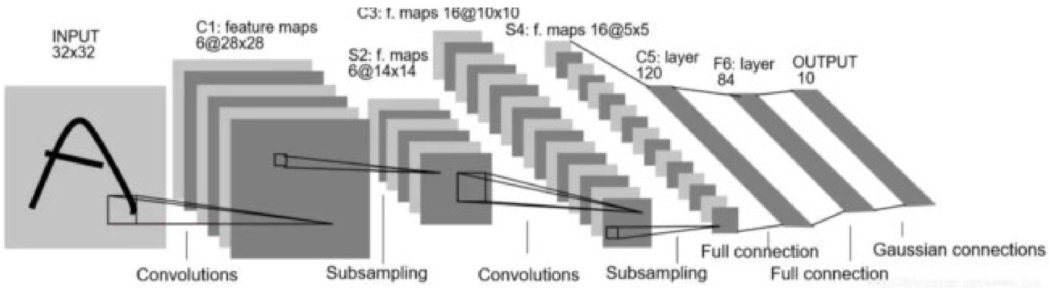

Figure 9. LeNet5

After LeNet5, due to the poor interpretability of neural network and the limitation of computing resources, the development of neural network in computer vision has been at a standstill until 2012. In 2012, Alex Krizhevsky, a student of Geoffrey Hinton, came up with AlexNet. AlexNet inherited LeNet5 idea and developed convolutional neural network into a very wide and deep network. AlexNet contains 630 million links, 60 million parameters and 650,000 neurons. Compared to LeNet5's 60,000 parameters, AlexNet is much more complex. The structure of AlexNet consists of five convolution layers, among which the first, second and fifth convolution layers are linked to the Maxpooling layer, followed by three fully connected layers.

4. Conclusion

As the development of human computer interaction progresses, several categories such as AI (artificial intelligence) are integrated into people's daily life. Programs such as Siri were programmed to perform psychological analysis on users, predicting the most favorable recommendations based on previously collected database. Human computer interactions have improved drastically along with various computer development. From the first computer being built in 1946 to personalization of computer for every household, our interactions with machines have increased considerably, and from direct manipulation through coding to connection of internet, we are able to collect and convey information faster and further than ever before. With the help of sizable database and developments of algorithms for sensory detection, computers are able to the interpret human interactions in a grander scale and more actively. With that being said, there are plenty of space for the development of human computer interactions in the bright future, when computers no longer solely respond to the environment, but actively engage with people and surroundings. The role for Human computer interaction was to let computers revolved around us in our daily life.

However, obstacles are also presented in the process of exploring human computer interactions, especially three major categories including the debates of morality, privacy, and safety.

Regarding moral issues, the concern is how to determine the rectitude of performances by the machine. For example, when artificial intelligence installed in a car is designed to avoid possible car crash, is it justifiable for the car to move out the way and potentially run into a pedestrian? In other words, are some human computer interactions helpful if our benefits are at the expense of hurting other human? The conflicting moral standards will be a challenging aspect as human-computer interaction serves more functions than ever before.

Privacy is also another dilemma that developers must ponder upon. Since computer needs data from our daily life to establish neural networks and therefore make the most suitable decisions for users, our information is being collected on daily basis. People will need to give up their privacy for an AI system to learn and adapt, but how willingly and to what degree should we sacrifice our privacy to work? The answer to this question can vary among individuals, and it is hard for computers to manifest a demarcation between being helpful and being intrusive.

Many developers are also posing safety concerns for Human computer interactions. Since programs and algorithms can potentially be hacked or attacked by others. For example, if AI such as Alexia is being hacked, a person's personal information will be entirely exposed to the hacker, causing more troubles and dangers than not having an AI. If the development of Human Computer interactions created more insecurities for customers, it will violate the purpose of creating Human Computer interactions in the first place.

Therefore, HCI has a hopeful future, but we should also tread carefully when perfecting computers and integrating them into our lives. Interdisciplinary experts ,such as lawyer, philosophers and computer scientists in general, will also be needed to answer questions regarding morality, safety and privacy. In the end, we are looking for the best option to maximize benefits from Human Computer interactions within a safe boundary.

References

[1] Elhai J D, Rozgonjuk D. Editorial overview: Cyberpsychology: reviews of research on the intersection between computer technology use and human behavior [J]. Current Opinion in Psychology, 2020, 36: iv-vii.

[2] Xing L, Chen W, Wang J, et al. Long Range MLRS Fire Control Model's Design Based on Human-computer Intersection Primitive [J]. Ship Electronic Engineering, 2013.

[3] Majid A . Human Olfaction at the Intersection of Language, Culture, and Biology[J]. Trends in Cognitive Sciences, 2020.

[4] Sawantdesai M. Human Computer Interaction based on Psychology and Interactive system design. 2014.

[5] Lai C C, Tsai W H. Human-Computer Interaction Using Eye-Glasses Shape Information by 3-D Computer Vision Techniques. 1999.

[6] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[7] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[8] [Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [J]. IEEE Computer Society, 2013.

[9] Related S, Related S. Gradient-based learning applied to document recognition [61].

[10] Technicolor T, Related S, Technicolor T, et al. ImageNet Classification with Deep Convolutional Neural Networks [50].

[11] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014.