Virtual Touch: Computer Vision Augmented Touch-Free Scene Exploration for the Blind or Visually Impaired

Xixuan Julie Liu
New York University Abu Dhabi
Abu Dhabi 129188, UAE
xl2200@nyu.edu

Yi Fang
New York University Abu Dhabi
Abu Dhabi 129188, UAE
yfang@nyu.edu

Abstract

The Blind or Visually Impaired (BVI) individuals use haptics much more frequently than the healthy-sighted in their everyday lives to locate objects and acquire object details. This consequently puts them at higher risk of contracting the virus through close contact during a pandemic crisis (e.g. COVID-19). Traditional canes only give the BVIs limited perceptive range. Our project develops a wearable solution named Virtual Touch to augment the BVI's perceptive power so they can perceive objects near and far in their surrounding environment in a touch-free manner and consequently carry out activities of daily living during pandemics more intuitively, safely, and independently. The Virtual Touch feature contains a camera with a novel point-based neural network TouchNet tailored for real-time blind-centered object detection, and a headphone telling the BVI the semantic labels. Through finger pointing, the BVI end user indicates where he or she is paying attention to relative to their egocentric coordinate system, based on which we build attention-driven spatial intelligence.

1. Introduction

1.1. Background

The World Health Organization's data show that there are 39 million blind individuals and 246 million individuals with low vision worldwide [49]. Prior to the arrival of the pandemic COVID-19, sight loss and low vision pose significant challenges for the Blind or Visually Impaired (BVI) individuals to carry out activities of daily living according to reports [45, 34, 36, 29, 62]. An April 2020 survey [12] conducted by American Foundation for the Blind with over 1,921 participants reveals that sudden pandemic crises like the ongoing COVID-19 profoundly magnified those barriers and obstacles in the BVIs' daily lives (Figure 1 left). BVIs have limited perceptive range so they use haptics much more frequently than the healthy-sighted in their everyday lives to locate objects and acquire object details. In non-pandemic times, BVIs compensated their visual deficiencies with alternatives to vision, such as working with a guide person to help with grocery shopping or exploring a new environment by touch. However, these compensatory methods have become risk-inducing or mentally stressing due to health and safety guidelines during pandemics, such as avoiding contact with object surfaces or maintaining social distance [2]. The shortened perceptive range of the BVI and the consequent elevated pandemic risks motivate us to propose our project. We aim to develop a wearable solution to augment the BVI's perceptive power so that they can perceive objects in their surrounding environment in a touch-free manner and carry out activities of daily living during pandemics more intuitively, safely, and independently. We name the proposed wearable solution Virtual Touch (Figure 1 right).

1.2. Needs of the BVI.

A variety of surveys [12, 60, 17, 6] have been conducted to offer insight into the specific needs of the BVI community during pandemics and potential ways to address them. Based on the outcomes of the surveys, we identified primary needs of the visually impaired during pandemic crises: (1) the BVIs need to safely explore their surrounding objects in a touch-free manner so to reduce contraction risks through close contact [20, 21, 60, 17, 15, 4, 7, 51]; (2) they should also be given more long range perspective power that the traditional canes do not offer and human assistants cannot offer during pandemic times [1, 27, 64, 14, 12]. BVIs should be able to actively and selectively explore objects present in the surrounding environment (e.g. quickly locate the door knob to open the door, identify and push the desired elevator button, look for the cashier for checkout). This requires more interaction than assistive technologies of First Person Vision [37, 28] based on head motion and gaze direction. In addition to support for safe and active exploration, the BVIs highly appreciate wearable solutions of compact size (e.g. a smartphone) that meet their needs

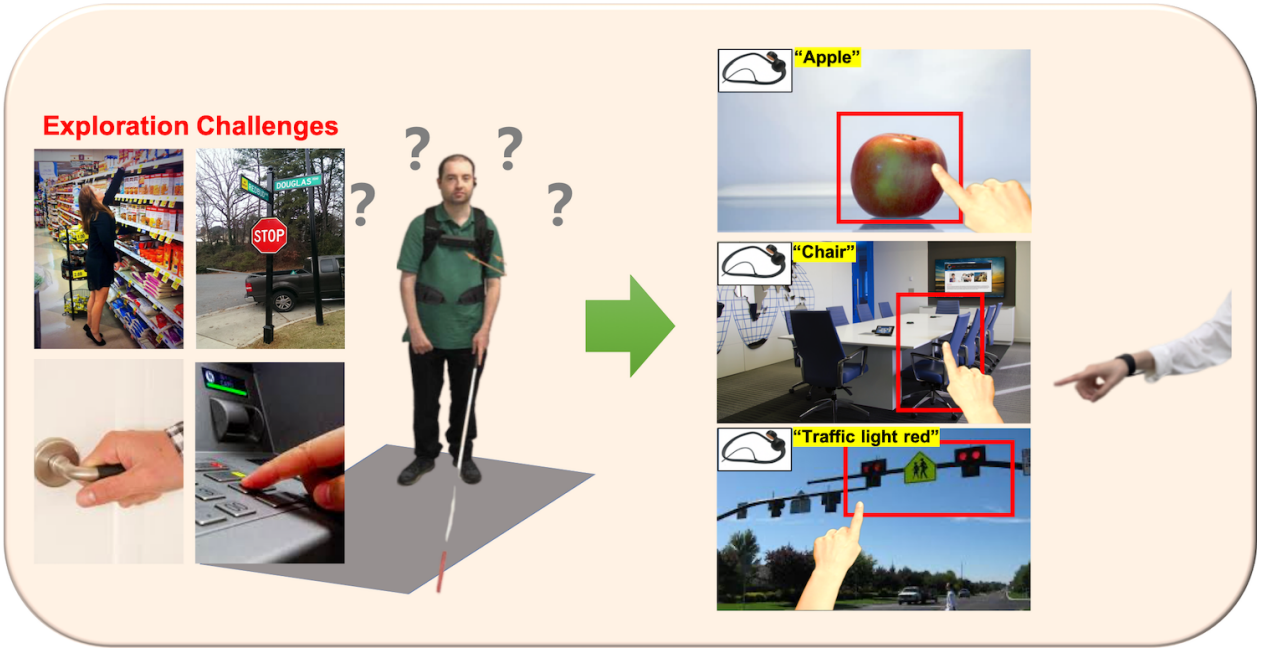

Figure 1. Exploration challenges for the BVI exist in various scenarios during the pandemics (left). The proposed Virtual Touch system replaces physical touch for the BVIs and assists them with safe environment exploration (right). [The images from Google search are copyright to their respective owners]

of mobility and functional independence while performing their activities of daily living [32, 48, 13, 53, 16].

1.3. Previous Assistive Technologies

Here we review the existing commercial assistive solutions for the BVI. The white cane [65] is a traditionally widely used [19, 68] and affordable tool for exploration of immediate surroundings, but its function is limited by its reach as a direct extension of the physical touch [45, 24, 57, 43, 69, 11, 56, 68]. Most hardware-based solutions face significant drawbacks that have led to an overall low adoption rate [25, 58, 44], including: (1) High cost. Popular commercial sensory substitution devices generally fall into a price range of a few thousand US dollars due to the high cost of dedicated hardware [5, 26, 3, 46, 31, 66]. (2) Cumbersomeness. As shown Figure 2, hardware-based assistive devices often lead to uncomfortable user experience due to the additional hardware components (i.e. sensors, battery). (3) Inability to address pandemic-specific needs. Most assistive technologies [58, 35, 40, 30, 39, 10, 33, 18, 9] are designed in a way that does not provide intuitive assistance under the context of viral pandemics. In contrast, software-based solutions that run a smartphone are more affordable and accessible for BVIs, with Microsoft Seeing AI and BlindSquare [47, 8] being two popular sensory substitution and navigation apps. Unfortunately, the assitive APPs technologies are unable to fully address the design requirements for risk reduction during a pandemic due to (1) lack of the functional design of an active mode that enables interactive communication between the user and surrounding environment, (2) lack of the functional design for the special pandemic needs (i.e. risk reduction), and (3) lack of on-board visual processing and dependency on online cloud visual computing (i.e. Microsoft Seeing AI), which could lead to functional failures.

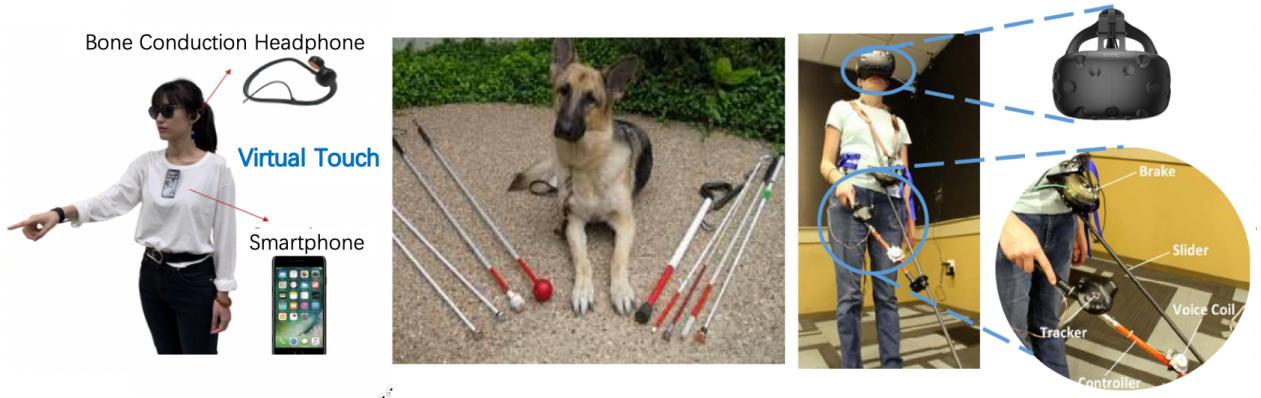

Figure 2. Existing assistive technologies for the BVI can be simple but functionally limited (middle) or well-equipped but cumbersome (right). The proposed Virtual Touch is lightweight, comfortable, and comprehensively assistive (left).

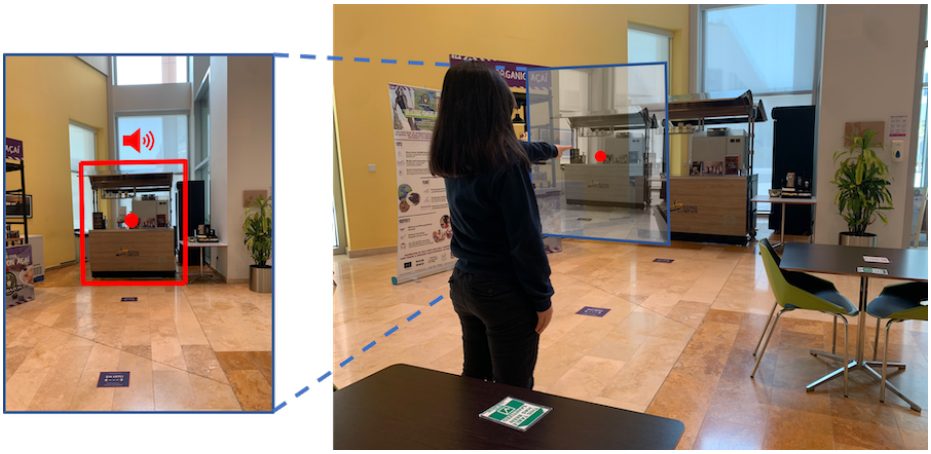

Figure 3. Virtual Touch helps the BVI touchlessly explore their surroundings. For instance, the BVI is notified of the coffee stand's presence when pointing in the coffee stand's direction.

1.4. Our Solution: Virtual Touch

Previous assistive technologies cannot meet the safety and health needs of the BVI in times of public health crisis to reduce physical touching on surfaces. They also fail in portability, interactivity and computational efficiency when trying to give the BVI augmented perception. We address these challenges by investigating ways to realize active environment exploration in a novel approach. Task specific algorithms are derived and used to extract key information via the phone's camera quickly and accurately, which enables real-time key object detection that improves scene understanding. Object recognition aided by the determination of user's focused attention will facilitate BVI's exploration. The work combines deep learning, computer vision, robotics, and rehabilitation engineering to bring breakthroughs to active mode AI assistance. The development of Virtual Touch transforms the assistive technology research from heavy, functionally limited devices or algorithms to a light-weight, interactive, and intuitive-to-use integrated system. Figure 3 illustrates Virtual Touch-assisted touch-free exploration of the environment.

2. Methods

2.1. Virtual Touch Hardware System

The Virtual Touch hardware system (dataflow shown in Figure 4) only requires a smartphone (with camera and GPU/CPU) and a pair of bone conduction headphones. The smartphone can be mounted at any location on the user such that it has an unobstructed frontal view of the surroundings and rests at a height roughly equal to the user's chest. By doing so, the smartphone has approximately the same view of the surroundings as one would see with their eyes from their head; at the same time, Virtual Touch can be mounted at a location that is comfortable. Virtual Touch is a Vision to Audio converter implemented through a smartphone application to relay the identification of a target object pointed to by a BVI person using audio through headphones.

2.1.1 Mobile Optimization

We leverage mobile-friendly computer vision techniques to ensure the core detection network TouchNet processes incoming information real-time. For the extraction of shared features, we utilized specialized convolutional neural network modules, namely inverted residual connections, depthwise separable convolutions, and use of longer strides [59, 67] in preference to pooling layer to decrease the computational workload while largely maintaining the representational power of the features extracted compared to large, workstation-oriented network architectures. The features fed to the predictor contain a combination of high-level and low-level semantic features about specific regions of the original image and is the foundation upon which subsequent visual processing takes place to accomplish classification and localization of objects.

2.1.2 Bone Conduction Headphone

The proposed system employs bone conduction headphones given its significant benefits to the wearer [63], including increased situation awareness and more comfortable fit. Traditional headphones and earbuds occupy the hearing ability of the BVI, and hearing is one the senses that they heavily rely on aside from touch. Considering that the BVI user is expected to wear the headphone for audio cues all the time when completing tasks, even when outdoors, the loss of awareness to natural sounds would cause inconvenience and pose an additional risk of accidents to the BVI. With bone conduction headphones, there are no speakers going over or into the ears, interfering with the BVI wearer's normal audition. Instead, the transducers sit on the cheekbones directly in front of the ears, leaving the ears completely open to the surroundings. Current studies on bone conduction have been confident about its performance even with 3D audio [42, 41].

2.2. Virtual Touch Software System

In this project, we use our deep learning model to directly classify the object being pointed at. The goal is to

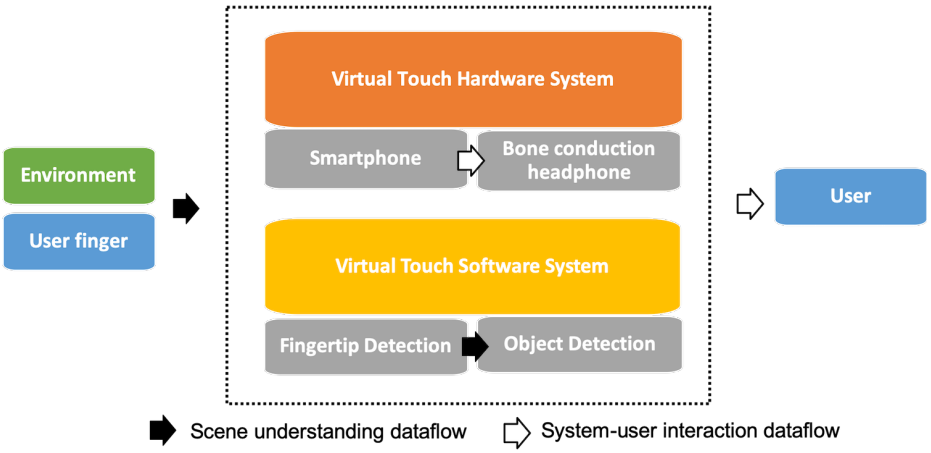

Figure 4. Closed loop dataflow for Virtual Touch.

have the model automatically focus on the object in the finger-pointed region. The network detecting the fingertip also leads to a region of natural attention, meaning the VI user's attention is mentally focused on the spatial direction around the fingertip. A neural network then extracts image features and detects objects only from the areas on feature maps that contain the finger tip location. The network has default anchor boxes of different aspect ratios at the specific locations corresponding to the finger tip location in several feature maps of different scales. For each default box, we predict both the shape offsets relative to the default box coordinates and the confidences for all object categories. Candidates exceeding a confidence threshold are taken as positions with objects. The predicted object with the highest confidence score is output as semantic notification to the user. The high level algorithm is described in Algorithm1.

2.2.1 Finger tip detection

Virtual Touch uses a histogram based approach to separate out the hand from the background frame. Thresholding and filtering techniques are used for background cancellation to obtain optimum results. To detect finger, the hand has to be differentiated from the background. A skin color histogram is used to subtracts the background from an image, only leaving parts of the image that contain skin tone [50, 52, 54]. Virtual Touch then finds the contour of the hand and determines the convexity defect, which is the furthest point from the centroid of the contour, as the tip of a finger. Another method to detect skin would be to find pixels that are in a certain RGB or HSV range. However, this approach would be sensitive to changing light conditions and skin colors. While on the other hand, our histogram approach tends to be more accurate and takes into account the current light conditions [50, 52, 54].

2.2.2 Attention driven object detection

Given a finger tip location, we feed to the network a clean image - the most recently preserved frame without user's hand detected. We use a predictor made of a classification head and a localization regression head to obtain all detected objects in the finger-pointed region before returning the one with the highest classification confidence. To this end, Figure 5 shows the procedural flow of our network. After the light and fast speed backbone network extracts features of different scales from the image, the predictor

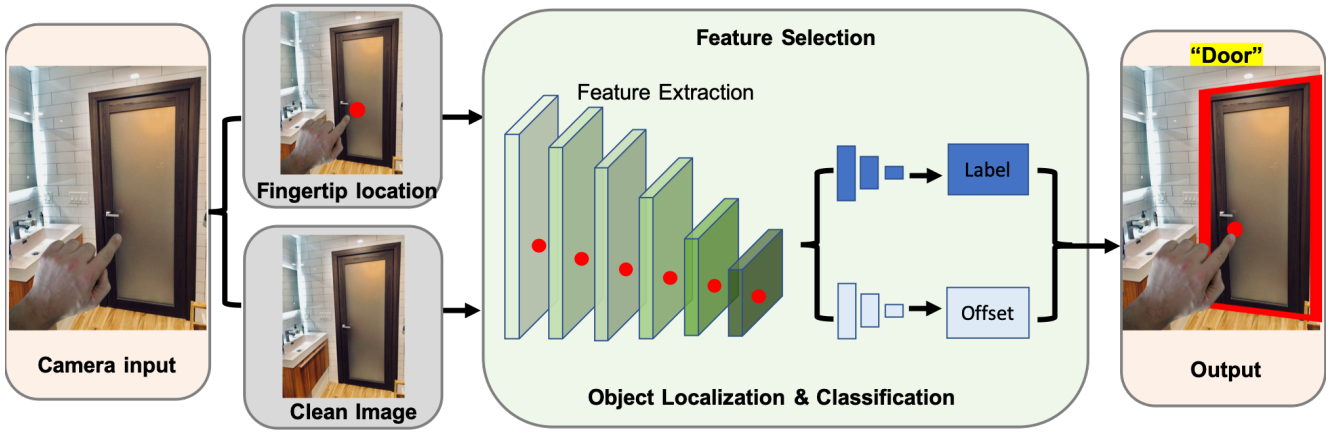

Figure 5. An illustration of point-based detection. [The images from Google search are copyright to their respective owners]

Algorithm 1: Algorithm of point-based detection

Data: Images fetched from camera
Result: Object detected given user's finger pointing
initialization of monocular camera, detection model, and TTS(Text to Speech Synthesizer) engine;
capture one frame from the camera;
keep it as the previous frame;
while *getting frame from the camera* **do**
  detect finger pointing at the current frame;
  if *no finger pointing is detected* **then**
    keep this clean image as the previous frame;
    continue;
  else
    get finger tip location;
    send the previous frame to model to detect objects of interest based on the finger tip location;
    if *objects are detected* **then**
      retain one detected object with the maximum confidence score;
      inform the user of the object information;
    else
      continue;
    end
  end
end

takes as input the concatenation of a number of features for classification and localization task. State of the art object detection algorithms [55, 38] commonly have anchors over the entire image and predictors scanning entire feature of each layer, which does not fit the goal of point-based detection. TouchNet, on the other hand, attaches anchor boxes only at areas that contain the finger pointed location and correspondingly only makes predictions looking at those areas on feature maps. This can cut off number of parameters to predict and computational cost. Thus, we take an performance-maximizing approach for the implementation of desired function at optimal accuracy with the minimum amount computational workload. The loss function we use is a combination of losses from the object localization (loc) and classification (cls) tasks.

$$L(x, c, l, g) = \frac{1}{N}(L_{cls}(x, c) + \alpha L_{loc}(x, l, g)) \quad (1)$$

where x is 1 if the prior is matched to the determined ground truth box, and 0 otherwise, N is the number of matched priors, l is predicted bounding box, g is ground-truth bounding box, c is class confidence. The bounding boxes are expressed by center offsets (cx, cy) and width-height (w, h). L_{cls} is classification loss, L_{loc} is localization loss, and α is hyper-parameter factor for balancing the weight of the losses determined by cross-validation. Smooth-L1 loss is used for localization on l and g, and Softmax loss is used for optimizing L_{cls} over multiple class confidences c.

3. Experiments

In this section, we carry out a set of experiments for point-based detection and assess the performance of our proposed TouchNet. In Sec. 3.1, we describe the details of datasets used for training and testing of TouchNet. In Sec. 3.2 we discuss the improvements of TouchNet on real time processing. In Sec. 3.3, we experimentally find out the best confidence threshold for TouchNet320 and report the performance of our trained TouchNet on standard test dataset. In Sec. 3.4, we demonstrate that our model can be used in real-world indoor and outdoor scenes.

3.1. Experimental Dataset

Our deep-neural-network model TouchNet for AI-enabled exploration is trained and tested on PASCAL Visual Object Classes Challenge (VOC) 2007 and 2012 [22, 23]. Each image contains a set of objects out of 20 different classes. The 20 classes are: Person - person; Animal - bird, cat, cow, dog, horse, sheep; Vehicle - aeroplane, bicycle, boat, bus, car, motorbike, train; Indoor - bottle, chair, dining table, potted plant, sofa, tv/monitor. Training of TouchNet used training set and validation set from both VOC 2007 and VOC 2012, which include 16,551 images in total. Evaluations in Sec. 3.2 and 3.3 are only based on VOC 2007 test set, which is 4,952 images. During training, all images went through data transformation to make the trained model more robust. Techniques used include photometric distortion (random contrast, color conversion, random saturation, and random hue), random expansion, random cropping, random mirroring, resize, and mean subtraction. During evaluation, images were only resized and mean-subtracted.

3.2. Detection inference time and computation analysis

In this experiment section, we demonstrated the improvement on inference time by our manipulations on the TouchNet model and analyzed the cut on computational costs.

Experimental Setting: To test inference time, we created one smaller network TouchNet320 and one bigger network TouchNet512. **TouchNet320** uses a backbone feature extractor of a fully convolutional layer with 32 filters, followed by 19 residual bottleneck layers [59]. The outputs for detection from the smaller backbone feature extractor are of size 20*20*96, 10*10*1280, 5*5*512, 3*3*256, 2*2*256, and 1*1*64. Each of the six feature maps is attached with 6 anchor boxes at the corresponding location of the finger

model	manipulation	FPS	input size
TouchNet512	Full feature detection + NMS	20	512*512
TouchNet512	Partial feature detection + NMS	21	512*512
TouchNet512	Partial feature detection + Confidence threshold	25	512*512
TouchNet320	Full feature detection + NMS	52	320*320
TouchNet320	Partial feature detection + NMS	54	320*320
TouchNet320	Partial feature detection + Confidence threshold	109	320*320

Table 1. Comparison of inference time of models on PASCAL VOC 2007.

tip. **TouchNet512** uses a bigger backbone feature extractor. It has 5 sets of 3 x 3 filters with stride of 2 in convolution layers and same padding in pooling layers 2 x 2 with stride of 2 [61], which are then followed by two additional 1 x 3 convolution layers with stride of 2 and padding of 1. The output features for detection are of size 64*64*512, 32*32*1024, 16*16*512, 8*8*256, 4*4*256, 2*2*256, and 1*1*256. Each feature map is attached with 4, 6, 6, 6, 6, 4, and 4 anchor boxes respectively, all at locations corresponding to the finger tip location. For each model, we tested full feature detection v.s. partial feature detection and NMS v.s. confidence threshold.

Result: Table 1 shows the increase in frame per second (FPS) or reduction of inference time by our design. Given that most of the inference time is spent on the backbone network [55, 38], using a faster base network could improve the speed significantly. Our experimental results confirm that observation as TouchNet320 has more FPS in general. We therefore use TouchNet320 as the deep learning neural network of Virtual Touch. On top of that, detecting objects only on part of features from all layers instead of entire features reduces the inference time at the predictor. Detecting objects based on user's attention simulates how humans perceive their surrounding environment in the most natural and cost-effective manner. Traditionally, considering the large number of boxes generated from all over the image, it is essential to perform non-maximum suppression (NMS) during inference. For example in SSD300, NMS with jaccard overlap of 0.45 per class and keeping the top 200 detections per image costs about 1.7 msec per image, which is close to the total time (2.4 msec) spent on all layers after backbone network [38]. In our model, there is no need for NMS. By using a confidence threshold, we filter out most unwanted boxes in a much faster way.

3.3. Detection performance analysis with pointed finger

In this section we evaluated the performance of our selected model TouchNet320. We designed a special evaluation metrics that fit the goal of AI-assisted environment exploration.

Experimental Setting: Here we tested the performance of TouchNet320 on 4,952 images in the VOC 2007 testset in order to determine the best confidence threshold. Each image was randomly assigned a point location which could be anywhere within the range of the image. We then defined Semantic Label Accuracy for evaluating the task-specific pointed-based detection model:

$$semantic\,label\,accuracy = \frac{true\ positive + true\ negative}{number\ of\ predictions} \quad (2)$$

where true positive is defined as correct label prediction of a object that contains the point or of the nearest object whose horizontal/vertical deviation from point less than half of image width/length. Different from the traditional metrics mAP, evaluation of TouchNet prediction does not have

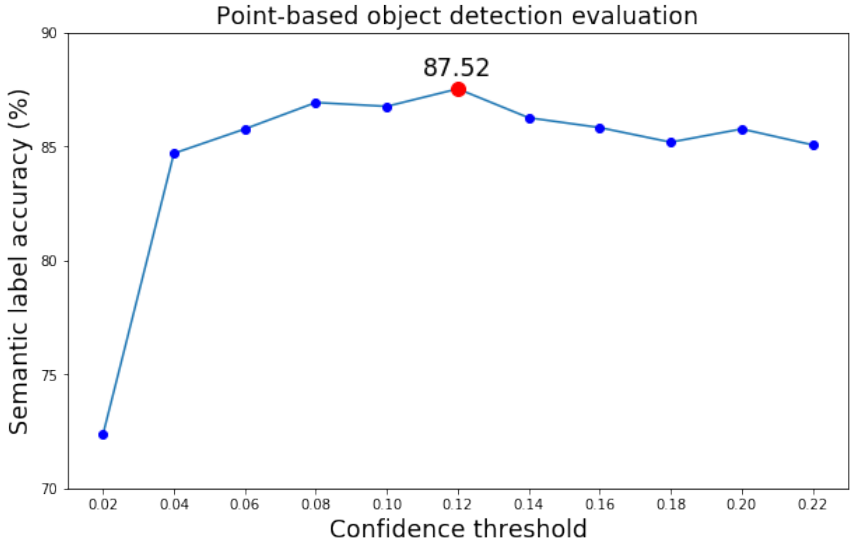

Figure 6. Comparison of performance of the TouchNet320 model on PASCAL VOC 2007, when using different confidence thresholds.

to worry about Intersection over Union (IoU) between predicted bounding box and ground truth box, as long as what the user's intended target object gets detected.

Result: Figure 6 shows the semantic label accuracies for the TouchNet320 network with different confidence thresholds. The best accuracy for the TouchNet320 network is 87.52% when the confidence threshold is 0.12, a performance value higher than the state of art mAPs on object detection [70]. We also conducted analysis on false negative predictions, which could assist future improvement. False negative could arise from a finger tip location on the edge of the target object. In this case, the part of feature maps containing the finger tip location and the part of feature maps containing the center of object might not be the same. Predictions made based on the location of finger tip is thus likely to be associated with a low confidence along with the true label of the object. Or it could be a hard-to-see object, for example, an object only partially present in the frame, or an object that is very far and small. This issue could be overcome with more layers in the backbone network, a larger input resolution for the network, or even more data augmentation during training.

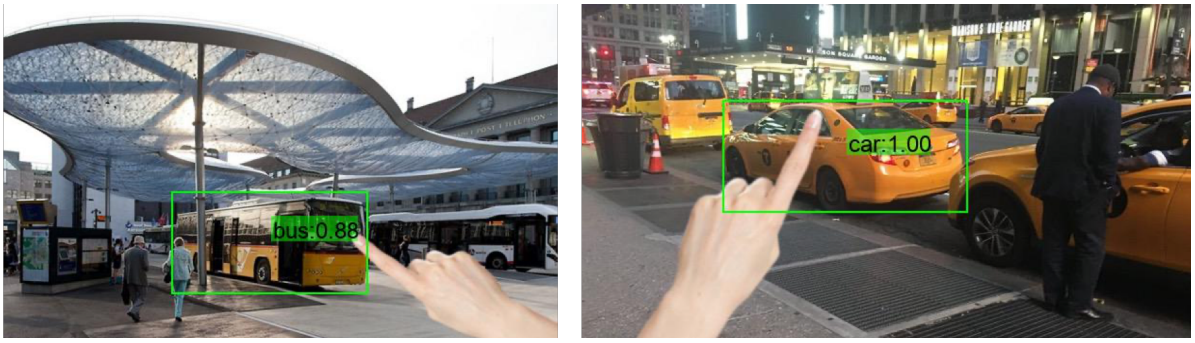

Figure 7. Examples of successful point-based detection in outdoor scenes. [The images from Google search are copyright to their respective owners]

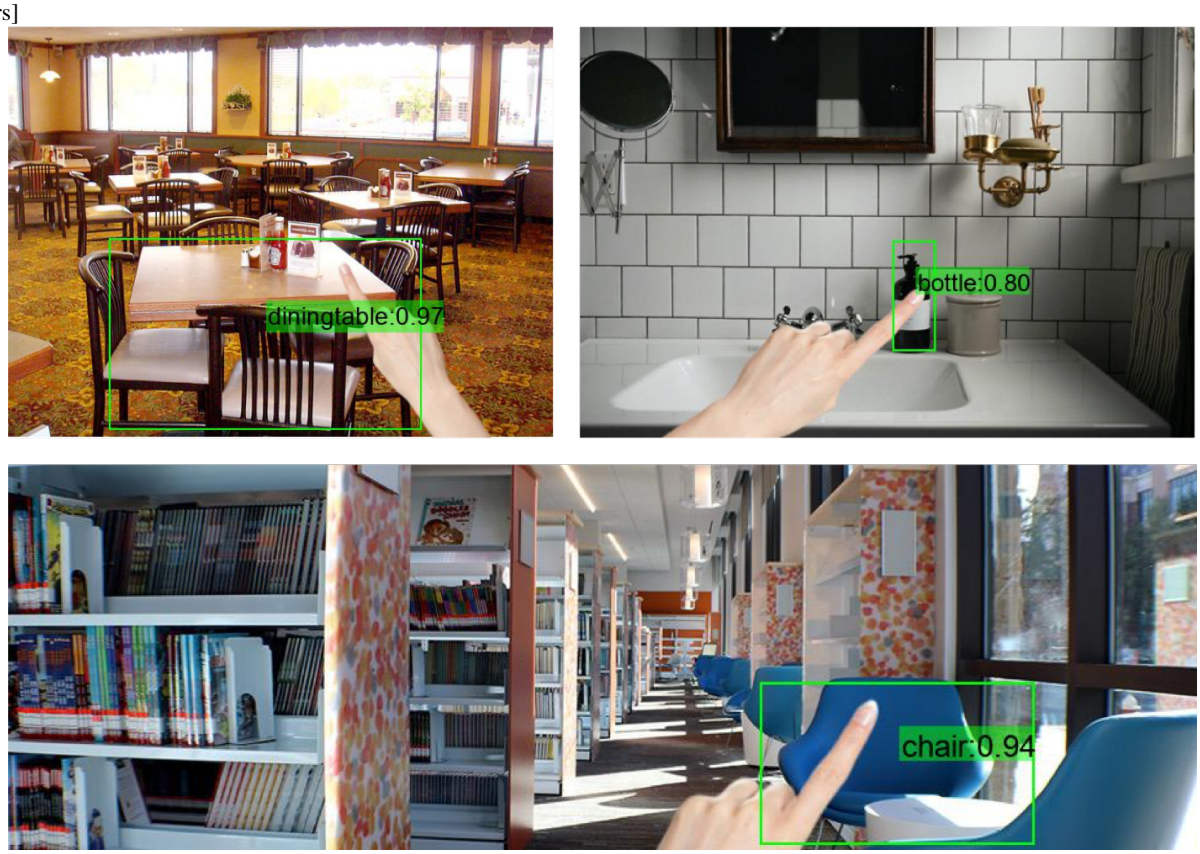

Figure 8. Examples of successful point-based detection in indoor scenes. [The images from Google search are copyright to their respective owners]

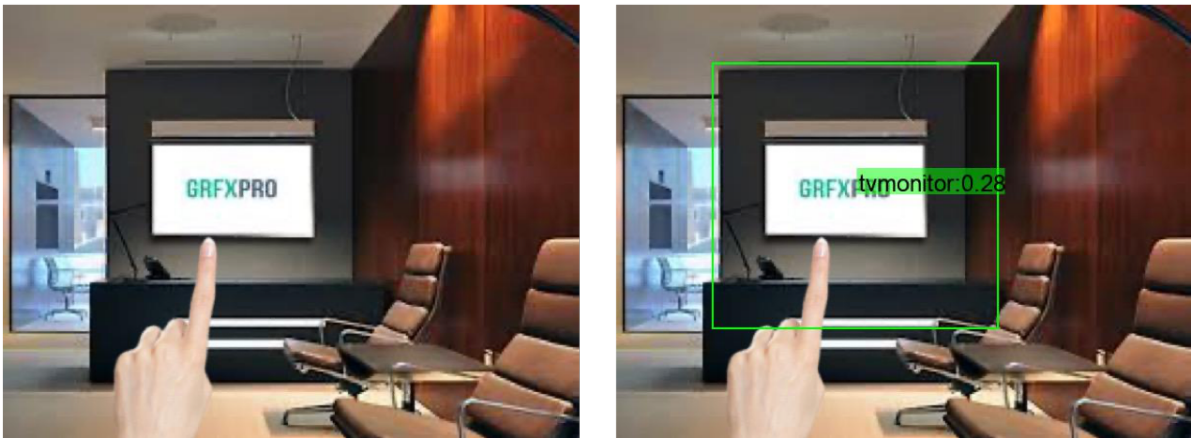

Figure 9. Example of detection result (TV monitor) dependent on backbone model. Left is false negative when using the TouchNet320 network; Right is true positive when using the TouchNet512 network. [The images from Google search are copyright to their respective owners]

3.4. Indoor and Outdoor Scene Detection Experiments

To evaluate our point-based detection model in actual indoor and outdoor scenes in real time, we tested it with our hands pointing at objects of interest in different settings.

Experimental Setting: To enable the video mode point-based detection, the user places their hand in the center of camera captured frames before using the Virtual Touch detection function. The finger tip detection algorithm takes skin color samples from the user's hand and then succeedingly extracts pixels from those frames to generate an HSV histogram. For the largest contour detected, it finds the hull, centroid, and defects. Among all these defects the application finds the one that is farthest from the center of the largest contour. This point location is kept as the location of pointing finger. The video mode takes frames at 30 FPS.

Result: When we tested video mode detection with a finger pointing to various objects, the objects being pointed at were correctly detected. We simulated some hands on scene images as examples for the purpose of display. The results in Figure 7 and Figure 8 show that Virtual Touch with its core TouchNet320 can successfully detect object at where the finger points to in real world settings. The objects detected can be in indoor scenes or outdoor scenes, from a very close distance to 10 meters away. Figure 9 shows that a heavier backbone network could capture more object. Especially when a larger input size is allowed, 512*512 instead of 320*320 for example, higher resolution leads to fewer detection failures, though it also means more processing time. Overall, our TouchNet320 based Virtual Touch has an optimal combination of speed and accuracy, and it has been proven useful in real life application.

4. Conclusions

In this paper, we present a novel assistive low-vision platform, Virtual Touch, that augments environmental understanding for the BVI while keeping them safe from contracting virus through touch-free exploration. We designed an attention based mechanism for intuitive indication of focus point using the position of fingertip. With that, we designed an end-to-end point-based neural network TouchNet to predict the locations and categories of pointed objects in real time. The experimental results demonstrated that this system can help the BVI understand their surroundings in an effective and efficient way. Therefore, the Virtual Touch assistive system backed by TouchNet addresses the design requirements for risk reduction during a pandemic.

Acknowledgement

The authors gratefully acknowledge the financial support from the NYUAD Institute (Research Enhancement Fund - RE132).

References

[1] Coming together during covid-19: Concerns and connections. *Coming Together During COVID-19: Concerns and Connections — National Federation of the Blind*, Mar 2020.

[2] Social distancing, quarantine, and isolation, May 2020.

[3] Wearable low vision glasses for visually impaired, May 2020.

[4] JADE ABDUL-MALIK. Living with vision loss during a coronavirus pandemic. 2020.

[5] UC Berkeley. Augmented reality for visually impaired people.

[6] Alexy Bhowmick and Shyamanta M Hazarika. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *Journal on Multimodal User Interfaces*, 11(2):149–172, 2017.

[7] SARAH BLAZONIS. 5 things to know: Covid-19 creates unique challenges for the blind and visually impaired. 2020.

[8] BlindSquare. Pioneering accessible navigation – indoors and outdoors, May 2020.

[9] Guido Bologna, Benoît Deville, Thierry Pun, and Michel Vinckenbosch. Transforming 3d coloured pixels into musical instrument notes for vision substitution applications. *EURASIP Journal on Image and Video Processing*, 2007:1–14, 2007.

[10] Brainport. Disabilities technology: Brainport technologies: United states.

[11] Anne Lesley Corn and Jane N Erin. *Foundations of low vision: Clinical and functional perspectives*. American Foundation for the Blind, 2010.

[12] Aira Tech Corp. Flattening the inaccessibility curve.

[13] James M Coughlan and Joshua Miele. Ar4vi: Ar as an accessibility tool for people with visual impairments. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 288–292. IEEE, 2017.

[14] Peterborough City Council. Peterborough association for the blind highlights impact of covid-19. *Peterborough City Council*.

[15] Michael Crossland. What coronavirus crisis means for blind and partially sighted people. 2020.

[16] Dimitrios Dakopoulos and Nikolaos G Bourbakis. Wearable obstacle avoidance electronic travel aids for blind: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):25–35, 2009.

[17] Zamir Dale. Experiences of deafblind persons during the covid-19 outbreak. 2020.

[18] David Charles Dewhurst. Audiotactile vision substitution system, Aug. 7 2012. US Patent 8,239,032.

[19] Bradley E Dougherty, K Bradley Kehler, Richard Jamara, Nicole Patterson, Denise Valenti, and Fuensanta A Vera-Diaz. Abandonment of low vision devices in an outpatient population. *Optometry and vision science: official publication of the American Academy of Optometry*, 88(11):1283, 2011.

[20] Devin Dwyer and Jacqueline Yoo. Facing coronavirus while deaf and blind: 'everything relies on touch'. *ABC News*, Apr 2020.

[21] Devin Dwyer and Jacqueline Yoo. Facing coronavirus while deaf and blind: 'everything relies on touch'. 2020.

[22] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[23] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[24] Nicholas A Giudice and Gordon E Legge. Blind navigation and the role of technology. *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*, 8:479–500, 2008.

[25] Nicholas A Giudice and Gordon E Legge. Blind navigation and the role of technology. 2008.

[26] Google. Google glass.

[27] Keith D. Gordon. survey: The impact of the covid-19 pandemic on canadians who are blind deaf-blind, and partially-sighted, 2020.

[28] Cazzato D;Leo M;Distante C;Voos H;. When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking.

[29] Sharon A Haymes, Alan W Johnston, and Anthony D Heyes. Relationship between vision impairment and ability to perform activities of daily living. *Ophthalmic and Physiological Optics*, 22(2):79–91, 2002.

[30] Abdelsalam Helal, Steven Edwin Moore, and Balaji Ramachandran. Drishti: An integrated navigation system for visually impaired and disabled. In *Proceedings fifth international symposium on wearable computers*, pages 149–156. IEEE, 2001.

[31] HTC. Htc vive.

[32] Alyssa Jackson. The hidden struggles america's disabled are facing during the coronavirus pandemic. 2020.

[33] Hiroyuki Kajimoto, Yonezo Kanno, and Susumu Tachi. Forehead electro-tactile display for vision substitution. In *Proc. EuroHaptics*, 2006.

[34] Gertrudis IJM Kempen, Judith Ballemans, Adelita V Ranchor, Ger HMB van Rens, and GA Rixt Zijlstra. The impact of low vision on activities of daily living, symptoms of depression, feelings of anxiety and social support in community-living older adults seeking vision rehabilitation services. *Quality of life research*, 21(8):1405–1411, 2012.

[35] Vladimir Kulyukin, Chaitanya Gharpure, John Nicholson, and Sachin Pavithran. Rfid in robot-assisted indoor navigation for the visually impaired. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 2, pages 1979–1984. IEEE, 2004.

[36] Ecosse L Lamoureux, Jennifer B Hassell, and Jill E Keeffe. The determinants of participation in activities of daily living in people with impaired vision. *American journal of ophthalmology*, 137(2):265–270, 2004.

[37] M. Leo, G. Medioni, M. Trivedi, T. Kanade, and G.M. Farinella. Computer vision for assistive technologies. *Computer Vision and Image Understanding*, 154:1–15, 2017.

[38] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. *Lecture Notes in Computer Science*, page 21–37, 2016.

[39] Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. Gps-based navigation systems for the visually impaired. *Fundamentals of wearable computers and augmented reality*, 429:46, 2001.

[40] Jack M Loomis, Reginald G Golledge, Roberta L Klatzky, Jon M Speigle, and Jerome Tietz. Personal guidance system for the visually impaired. In *Proceedings of the first annual ACM conference on Assistive technologies*, pages 85–91, 1994.

[41] Raymond M. Measurement and validation of bone-conduction adjustment functions in virtual 3d audio displays. *SMARTech Home*, Jul 2009.

[42] Justin A. MacDonald, Paula P. Henry, and Tomasz R. Letowski. Spatial audio through a bone conduction interface. *International Journal of Audiology*, 2006.

[43] Shachar Maidenbaum, Sami Abboud, and Amir Amedi. Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews*, 41:3–15, 2014.

[44] Shachar Maidenbaum, Sami Abboud, and Amir Amedi. Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews*, 41:3–15, 2014.

[45] Daniela Massiceti, Stephen Lloyd Hicks, and Joram Jacob van Rheede. Stereosonic vision: Exploring visual-to-auditory sensory substitution mappings in an immersive virtual reality navigation paradigm. *PloS one*, 13(7):e0199389, 2018.

[46] Microsoft. Microsoft hololens.

[47] Microsoft. Seeing ai app from microsoft.

[48] Patrick Emeka Okonji and Darlinton Chukwunalu Ogwezzy. Awareness and barriers to adoption of assistive technologies among visually impaired people in nigeria. *Assistive Technology*, 31(4):209–219, 2019.

[49] W.H. Organization. Visual impairment and blindness, 2010.

[50] Amar Prakash Pandey. Finger detection and tracking using opencv and python. *DEV Community*, Sep 2019.

[51] Katherine Phillips. Coronavirus pandemic causes unique challenges for visually impaired, those with special needs. 2020.

[52] Rosalyn Porle, Ali Chekima, F. Wong, and G. Sainarayanan. Performance of histogram-based skin colour segmentation for arms detection in human motion analysis application. *International Journal of Electronics, Communications and Computer Engineering*, 1:403–408, 01 2009.

[53] Shi Qiu, Ting Han, Hirotaka Osawa, Matthias Rauterberg, and Jun Hu. Hci design for people with visual disability in social interaction. In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pages 124–134. Springer, 2018.

[54] J. L. Raheja, Karen Das, and Ankit Chaudhary. Fingertip detection: A fast method with natural hand. 2012.

[55] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.

[56] John-Ross Rizzo, Kyle Conti, Teena Thomas, Todd E Hudson, Robert Wall Emerson, and Dae Shik Kim. A new primary mobility tool for the visually impaired: A white cane—adaptive mobility device hybrid. *Assistive Technology*, 30(5):219–225, 2018.

[57] Uta R Roentgen, Gert Jan Gelderblom, Mathijs Soede, and Luc P De Witte. Inventory of electronic mobility aids for persons with visual impairments: A literature review. *Journal of Visual Impairment & Blindness*, 102(11):702–724, 2008.

[58] Uta R Roentgen, Gert Jan Gelderblom, Mathijs Soede, and Luc P De Witte. Inventory of electronic mobility aids for persons with visual impairments: A literature review. *Journal of Visual Impairment & Blindness*, 102(11):702–724, 2008.

[59] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2018.

[60] Rick Yarborough Scott MacFarlane and Jeff Piper. Blind community faces shopping challenges during pandemic. 2020.

[61] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. 2014.

[62] Joan Stelmack. Quality of life of low-vision patients and outcomes of low-vision rehabilitation. *Optometry and Vision Science*, 78(5):335–342, 2001.

[63] Mary Tang. Benefits of bone conduction and bone conduction headphones. *Medium*, Jul 2019.

[64] Steve Topple, Tracy Keeling, The Canary, and John Ranson. Blind and partially sighted people seem to have been forgotten in coronavirus policy. *The Canary*, Apr 2020.

[65] Troy L. McDaniel. Haptic belt, 2019. [Online; accessed 16-July-2019].

[66] Valve. Valve index.

[67] Robert J. Wang, Xiang Li, and Charles X. Ling. Pelee: A real-time object detection system on mobile devices, 2018.

[68] Gale R Watson, WILLIAM DE L'AUNE, Joan Stelmack, Joseph Maino, Sharon Long, and Gale R Watson. National survey of the impact of low vision device use among veterans. *Optometry and vision science*, 74(5):249–259, 1997.

[69] Wikipedia contributors. White cane — Wikipedia, the free encyclopedia, 2019. [Online; accessed 16-July-2019].

[70] Papers with Code. Pascal voc 2007 leaderboard.