

Mapping Out Paths for the Blind With Machine Learning

Originally published at <https://www.unite.ai/mapping-out-paths-for-the-blind-with-machine-learning/>



New research out of Germany offers a novel, GPU-powered portable system to help vision-impaired people to navigate in the real world. The system addresses one of the core challenges in real-time computer vision frameworks – the identification of glass and other transparent obstacles.

The [paper](#), from the Karlsruhe Institute of Technology, details the construction of a user-worn system, entitled Trans4Trans, consisting of a pair of smart glasses connected to portable GPU casing, effectively a lightweight laptop, which captures RGB and depth images at 640×480 pixels in a continuous stream, which is then run through a semantic segmentation framework.



The mobile sensors in the Trans4Trans rig. Source: <https://arxiv.org/pdf/2107.03172.pdf>

The system's sensory feedback capabilities are boosted by a pair of bone-conducting earphones, which emit acoustic feedback in response to environmental obstacles.

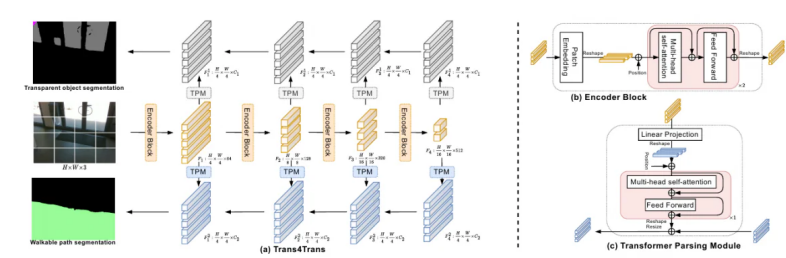
The Trans4Trans system has also been tested on the Microsoft HoloLens 2 augmented reality rig, achieving complete and consistent segmentation (i.e. recognition) of potentially dangerous obstructions such as glass doors.



Trans4Trans running on the HoloLens 2.

Architecture

Trans4Trans uses a dual approach, utilizing both a [transformer-based](#) encoder and a decoder, and leveraging a proprietary *Transformer Pairing Module* (TPM) capable of collating feature maps generated by the embeddings of dense partitions, while the transformer-based decoder is able to consistently parse feature maps from its paired encoder.



The architecture of Trans4Trans.

Each TPM consists of a single transformer-based layer, essential for the low resource drain and portability of the system. The decoder contains four symmetrical stages for the encoder, with a TPM module assigned to each. The system makes resource savings by integrating the functionality of multiple approaches into a coherent system, instead of deploying two separate models in a linear work-flow.

Hardware

The glasses used in the system incorporate a RealSense R200 RGB-D sensor, while the host machine houses a [Jetson AGX Xavier](#) NVIDIA GPU, designed for embedded systems, and featuring 384 NVIDIA CUDA cores and 48 Tensor cores.

The R200 offers [speckle projecting](#) and passive stereo matching, making it suitable for interior and exterior environments. The speckling facility is of particular benefit in evaluating transparent surfaces, since it augments and clarifies the incoming visual data without becoming blinded by extreme light sources. The sensor's infrared capabilities also help to obtain distinct geometry and form actionable depth maps, which are critical for obstacle avoidance, in the context of the aims of the project.

Preventing Cognitive Overload for the User

The system needs to strike a balance between adequate data frequency and excessive information, since the wearer needs to be able to distinguish the environment coherently through audio feedback and vibration feedback.

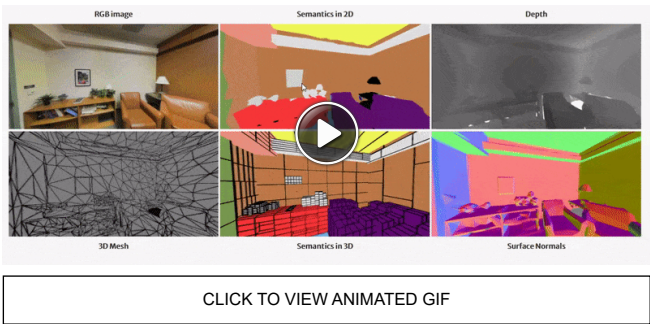
Consequently Trans4Trans artificially limits the volume of feedback data, with a single default threshold set to one meter, rather than forcing the user to learn a variety of vibration settings that accord with varying distances of looming objects and barriers.

Testing Trans4Trans

The Trans4Trans system was tested on two datasets dealing with the segmentation of transparent objects: [Trans10K-V2](#), from the University of Hong Kong *et al*, which contains 10,428 images of transparent objects for validation, training and testing; and the Stanford2D3D dataset, which contains 70,496 images of mixed transparency objects, captured at 1080×1080 resolution.



Images and corresponding masks from the Trans10k dataset. Source: <https://arxiv.org/pdf/2101.08461.pdf>



The Stanford2D3D system in action. Source: <http://buildingparser.stanford.edu/dataset.html>

In testing, Trans4Trans was also able to segment transparent objects that were misclassified by the [Trans2Seg initiative](#) released at the start of 2021 by the same researchers, while requiring fewer GFLOPS to calculate and segment the surfaces.

Unlike Trans2Seq, which utilizes a CNN-based encoder and transformer-based decoder, Trans4Trans uses only transformer-based encoder-decoder architecture, outperforming the previous approach and also greatly improving on PVT.

The algorithm also achieved state-of-the-art results for a particular number of transparent classes, including *jar*, *window*, *door*, *cup*, *box* and *bottle*.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More