Please list 3 circumstances that this system will fail.

Firstly, the system will fail when environmental conditions are not sufficiently similar to that during the training data collections phase, i.e. not bright and sunny. This would include times when it is raining heavily or during the night. This is because the object detection algorithm could be thrown off due to the different lighting or distractions (like water droplets), thus it is unable to locate the bounding box that contains the bus number.

Secondly, the system will fail when the user does not orientate the camera properly to view the bus. For the system to detect the bus number on buses, the camera has to be orientated such that it faces toward the oncoming buses. If the visually impaired user does not orientate the camera properly, for example, if he looks directly at the road instead of towards the oncoming traffic, the bus number could be hidden in the image or extremely skewed such that it becomes unrecognizable.

Thirdly, the system will fail when there are multiple different buses in the same image. This can happen sometimes when multiple buses arrive at the same bus stop. In this case, while it may be preferable for the system to mention all the buses that arrive (in chronological order), the algorithm may be locked onto one bus which it has a less obstructed view of the bus number. For example, the algorithm may lock onto the front bus amongst a pair of buses arriving simultaneously as the first bus’s bus number is larger and thus the algorithm has a higher confidence of identifying and thus only reads of the bus number of the first bus.

Provide Solutions to each of the circumstances listed in question 1

For the first situation, we could improve this with improved training data. Instead of merely training the algorithm on images in sunny day environments, we could increase the variety used in the training of the object detection algorithm to include many different locations and environmental conditions to allow the algorithm to develop a “more general” understanding of how to locate the bus number on a bus in an image.

For the second situation, we could include a supplementary system to aid the user in orientating himself correctly. This could come in the form of a classification neural network that takes in the view of the camera and has a single value, the angle by which the camera should be rotated to be correctly oriented. For example, the last node could have an arc tangent activation function with a large negative output indicating the need to turn clockwise, a near zero output indicating that the camera is correctly oriented and a large positive value indicating the need to turn counter-clockwise. The magnitude of the turning angle is determined by the magnitude of the output. This would provide information to the user to help the user quickly orientate the camera properly for the system to be effective.

For the last situation, we could improve the system with object tracking to monitor how the bus numbers (and thus buses) are moving from one frame to another. We could then code the system such that after it has mentioned a bus number of a bus, it will ignore all subsequent bounding boxes of that bus. Thus, it will not be locked on to buses that it has already read the bus number of, instead, it will go for other bus numbers that are visible in the image, thus ensuring that the user does not miss out his bus due to multiple buses arriving at the same bus stop at once.

How to integrate the solutions such that they do not “fight” over what the correct output is?

In terms of integration of these solutions, the solutions mentioned above do not conflict with one another but rather add-on to the system independently. However, it is noted that it may be possible that training the object detection algorithm with different environment types ends up diluting its specialization at the initial “bright day” environment. If this is the case, accuracy can be improved by training different object detection algorithms with images of buses in different environments and use a simple classification neural network to identify what type of environment the image is in and thus which of the object detection algorithms to use. This could even be combined together with the neural network from the second solution such as to reduce the overhead in calling multiple neural networks instead of just calling a single neural network.