Pedestrian Assistant: Computer Vision to Help the Visually Impaired

LITERATURE REVIEW

DARSHAN KRISHNASWAMY

`

Using Computer Vision to help blind pedestrians navigate safely.

Literature Review

Introduction

Visually impaired pedestrians are generally at a high risk for getting into accidents because they lack the ability to see cars or other obstructions in front of them. Between 2007 and 2016, total pedestrian deaths increased by 27 percent. In 2017, the estimated number of pedestrian fatalities was close to 6,000 (Retting, 2017). In comparison, there were closer to 4,000 such deaths in 2009, with 24 percent of these accidents occurring at intersections (Pharr, Coughenhour, Bungum, 2013). These statistics demonstrate that although precautions are in place to protect pedestrians, they are still ineffective at preventing many deaths. Such accidents are especially common among blind and visually impaired pedestrians. In a survey conducted on visually impaired pedestrians by the American Council of the Blind, 8 percent of respondents reported that they had been hit by a car at an intersection, and 28 percent of respondents reported that their canes had been run over (Al-Fuqaha, Oh, Kwigizile, 2018). Visually impaired pedestrians have such a high rate of accidents due to their decreased ability to detect pedestrian traffic lights and obstructions that may be in the way,such as people, poles, and street signs. Currently, audible signals are used at some pedestrian traffic lights to assist the visually impaired; however, this system comes with many problems and inconsistencies. A study done in 2012 demonstrated that blind pedestrians performed significantly worse than the normally sighted, even with access to auditory information (Hassan, Massof, 2012).

Pedestrian Crossings

Two common types of pedestrian crossings in the United States are the zebra crossing and the signal intersection. The zebra crossing consists of striped lines on a street where pedestrians may safely cross. The second type of crossing is the type that involves a special traffic signal designated for pedestrians, which are referred to as signal intersections. Pedestrians may cross zebra crossings at any time, as they always have the right of way at these crossings, and cars are required to yield to anyone waiting to cross at these intersections. At signal intersections, pedestrians only have the right of way when their light is green. These types of crossings are generally safer than zebra crossings, since the pedestrians only have a green light when cars have a red light. At zebra crossings, however, cars may still be traveling while the pedestrian is trying to cross, which can be dangerous (Federal Highway Administration, 1999).

Jaywalking

One especially common cause of pedestrian injuries is jaywalking Jaywalking refers to any time that a pedestrian walks into the path of a motor vehicle when he or she does not have the right of way. Jaywalking can occur either at signal intersections, when the pedestrian light is red or at other points along the road that are not designated pedestrian crossings. Jaywalking is the cause of about 6,000 driving accidents every year (Team, 2016).

Blind pedestrians often jaywalk on accident, since they may be unaware that they are at an intersection or that the pedestrian traffic light is red. Because of this, many states and towns have White Cane Laws, which require all cars to yield their right of way to any blind pedestrian with either a white cane or a guide dog (Murphy, 1965). Although these laws help to protect blind pedestrians who are unaware of their surroundings, the system is not perfect, and, as mentioned earlier, there are still cases every year where blind people are hit by moving cars while the blind pedestrians are supposed to have the right of way (Al-Fuqaha, Oh, Kwigizile, 2018).

Pedestrian Traffic Lights

Most common in large cities, some intersections contain special traffic lights for pedestrians. These are mainly used in areas where drivers may find it difficult to see pedestrians, as the drivers do not need to actively search for pedestrians and stop for them. Instead, pedestrians can cross at designated times and are required to wait during other times. Unlike normal traffic lights, pedestrian traffic lights can vary between different locations. The standard pedestrian traffic light displays a solid red hand when it is not safe to cross, a walking person when it is safe to cross, and a blinking red hand when it is not safe to begin crossing, but safe to continue crossing if the person has already started walking. However, some pedestrian lights appear differently. For example, some display a red person standing when it is not safe to cross, and a green person walking when it is safe to cross. Others display the words “WALK” in white and “DONT WALK” in red when it is safe and not safe to cross, respectively. **Figure 1** displays the different types of pedestrian traffic lights.

Figure 1. The three common types of pedestrian traffic light. Retrieved from ("New York 'Walk / Don't Walk' Sign", n,d,), (Vani, 2017), ("300mm Green Walk Man Pedestrian Crossing Light For Road Safety", n.d.)

Audible Pedestrian Signals

Currently, the most common way for pedestrian traffic signals to be accessible for pedestrians is the use of audible signals that play during the WALK interval. They generally play a chirping sound, similar to the sound that birds make, and they have distinct sounds for North-South and East-West intervals, which allow visually impaired pedestrians to know which direction has the green light (Szeto, Valerio, Novak, 1991). Not all audible pedestrian signals play the same sound. Along with the chirping sound, which is commonly referred to as a "cuckoo" sound, audible signals may also play sounds such as a "rapid tick" sound, where a ticking noise is played multiple times per second. Another possible sound is a noise referred to as a "peep-peep noise." These audible signals are not legally required at pedestrian intersections. Instead, they are often added as requested by residents of the city. In the city of Columbus, for example, only about 10.5 percent of pedestrian crossings with traffic signals have audible crosswalk signals (Kistler, 2015).

While audible signals generally help visually impaired pedestrians know when it is safe to cross the street, the system is not perfect. At times, pedestrians may mistake the sound of a real bird chirping for the audible pedestrian signal and think it is safe to cross. The different tones that play at different intersections also often confuse pedestrians. Pedestrians also often have trouble remembering which noise corresponded with each direction, and also sometimes did not know which direction they were facing (Kistler, 2015).

Machine Learning

One possible way to resolve the issue of allowing signals to be accessible to visually impaired people is through a system that uses machine learning to assist the user. Machine learning is a way of creating a model that is able to learn from data. There are three major types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the model is given data and corresponding "labels," which are the desired output. The model is trained to take in data and return the correct output. In unsupervised learning, there are no labels; instead, the model looks for patterns in the data that it is given and learns how to sort the data. Reinforcement learning is mainly used in robotics and game development. It works by giving the model a "reward" for a successful action and the model learns how to maximize its reward (Thrun, 1992). A potential solution to the pedestrian traffic light problem for visually impaired pedestrians would be to use supervised learning to create a model that would be able to detect pedestrian traffic lights and alert the user of the light's status (whether it is red or green).

Neural Networks

In order to classify the different types of pedestrian traffic lights, a neural network must be created to perform the classification. A neural network is a machine learning model. Neural networks are comprised of layers, each connected to the previous layer. The first layer of a neural network is the input layer, which contains the original data. Each value in the next layer is produced by taking each value in the previous layer, multiplying each value by a distinct number, known as a weight, adding up all the resultant values, and adding another term known as a bias. The values of this layer are then passed to the next layer, and the process continues throughout all of the layers until the final layer, known as the output layer.

Layers are comprised of neurons, each holding a single number. The number of neurons in a layer generally decreases as one approaches the output layer. There are many possibilities for the number of neurons in the output layer. If the model's purpose is to predict a value, for example, the output layer could have a single neuron containing its prediction (Lapedes, Farber, 1988). If the model's purpose is to classify an image between 10 categories, it could have 10 neurons in the final layer, each containing the probability that it is an image of each category. **Figure 2** illustrates how the values in one layer of a neural network correspond to the values in the next layer.

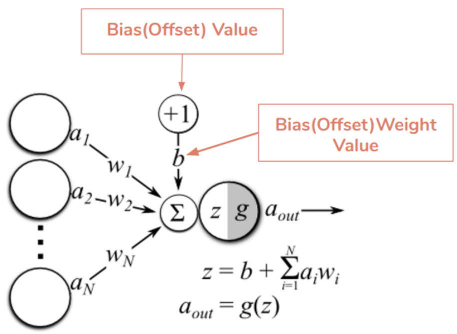

Figure 2. A visual representation of how a neural network layer works. Retrieved from ("Everything you need to know about Neural Networks", 2017).

Activation and Loss Functions

There are two important functions used by neural networks: the activation function and the loss function. The activation function is applied to each neuron, and it can serve many different purposes. One major purpose of the activation function is to create an upper and lower bound for the neurons' values. Activation functions such as the sigmoid and hyperbolic tangent functions, two common activation functions which are based on the exponential function, have fixed output ranges, which prevents the neurons' values from getting too large or too small. Another purpose of an activation function is to eliminate the linearity of the model. Without an activation function, each layer would be calculated as a linear function of the previous layer, so the entire neural network would just be a linear function of the input layer. Activation functions, which are usually nonlinear functions, prevent this and increase the complexity of the model, allowing it to reach a higher accuracy than it would without an activation function (Sharma, 2017). The loss function represents the neural network's error. There are many different loss functions. One is called mean-squared error, in which the difference between the values in the output layer are subtracted from the truth values, those values are squared, and the average of the resultant values is calculated. Other loss functions include softmax cross-entropy, which is commonly used in classification, and mean-absolute error, which is similar to mean-squared error, but instead of squaring the values, the absolute value is taken (Parmar, 2018). The loss function is used in gradient descent, a process used to optimize the parameters which will be discussed further in the next section.

Gradient Descent

Once the traffic light neural network has been initialized with the appropriate number of layers and size of each layer, the model must be trained in order to have an accurate output. When a neural network is initialized, all of its parameters (weights and biases) are randomly initialized. The neural network then needs to go through a training process, in which these parameters are tuned so that they create accurate outputs. This process of training a neural network is made possible through what is known as gradient descent. The process involves finding the gradient vector of the loss function, and then adjusting all of the parameters by subtracting the gradient multiplied by another parameter known as the learning rate, which must be manually inputted before training. The gradient vector points in the direction of greatest ascent (increase) for the loss function, so the negative gradient points in the direction of steepest descent. Thus, by subtracting the gradient multiplied by some value, the model decreases the value of the loss function, making the model more accurate (Pandey, 2019). Eventually, the loss function will taper off at some value, at which the accuracy of the model is maximized, and can only be affected by tuning other parameters such as the number of layers, the size of each layer, and the amount of data used to train the model. The python libraries Tensorflow, Pytorch, and Keras are commonly used for machine learning projects because they have optimized gradient calculation abilities, which are useful when training neural network models.

Convolutional Neural Networks (CNNs)

The solution to the pedestrian traffic light situation using neural networks is through image classification, the process of classifying an image as a member of a certain category. For example, in the pedestrian traffic light case, the categories would be red light and green light. Convolutional neural networks (CNNs) are a specific type of neural network that are commonly used for image processing and classification. The input image is a matrix with three dimensions: one dimension for the width of the image, one dimension for the height of the image, and one dimension for the image's colors (this dimension has a size of three, one for each of the colors red, green, and blue, which are known as the color channels). This matrix can be imagined as three two-dimensional matrices stacked on top of each other, one for each color channel. CNNs have two special types of layers: convolutional layers and max pooling layers. In convolutional layers, the CNN uses a window, known as a kernel, which is a two-dimensional matrix. This matrix "slides" over the image matrix, one color channel at a time, and each value of the image matrix is multiplied by the value in the kernel to obtain the layer's output value. In a max pooling layer, a new kernel slides over the image. This kernel contains no values; as it slides over the matrix, the maximum value covered by the kernel replaces the covered area (the output from this layer is usually smaller than the input). This process repeats multiple times, and the output is a classification layer, which generally contains one neuron for each category that the neural network is classifying between. For example, if a neural network was classifying images between dogs, televisions, and shoes, there would be three neurons in the final layer, each containing the probability that the input is an image of each respective category. The predicted category is the category with the highest probability from the output layer (Ciresan, Meier, Masci, Gambardella, Schmidhuber, 2011).

CNN Activation and Loss Functions

An important step in the process of creating the image classification model is choosing appropriate activation and loss functions. The most commonly used activation and loss function for CNNs are softmax and cross-entropy, respectively. Softmax is a function which takes in a set of values for each category and returns the probability that the image should be classified as each category. The cross-entropy function is a loss function that is very similar to mean absolute or mean squared error, in that it also calculates the distance between the output value and the expected value. It utilizes logarithms, which allows it to pair well with the softmax function, which uses exponential functions. Because of this, the two functions are frequently used together in convolutional neural networks (Liu, Wen, Yu, Yang, 2016).

Datasets and Partitioning

Neural networks "learn" how to classify between different categories by looking at data and finding patterns that allow it to differentiate between the categories. In image classification problems, an example of supervised learning, the data is accompanied by corresponding "truth values," which represent the labels for each image. For the pedestrian light neural network, this dataset will contain images of different types of pedestrian traffic lights, with labels containing the status (red or green) of each light.

As neural networks are trained on a set of data, they will eventually learn to find patterns that are specific to the data that they are trained on, which will cause it to have high accuracy in classifying the data that it has already seen, but low accuracy on data that it has not yet seen, a phenomenon known as "overfitting." To prevent this, the dataset is often partitioned into two sets: the train set and the validation set. The train set is the data that the model sees and is trained on, while the validation set is used to obtain an unbiased value of the accuracy of the model and to prevent overfitting the model (Moore, 2001). This is especially important in the pedestrian light model, as it will be classifying images that it has never seen before, so it needs to have high accuracy in predicting the status of images that it has not been trained on.

Tensorflow and Android Implementation

The portability of cellular devices allows them to be a convenient way for users to run machine learning projects, such as the pedestrian traffic light neural network. Applications can be created for Android devices using the Android Studio Development Environment, a free software created by Google. Android applications are created using the Java and Extensible Markup Language (XML) languages. While many other Android application development programs exist, Android Studio is the official development environment for Android devices and is the most commonly used due to its versatility and convenient features such as built-in device emulators (Technoaddict, 2018).

A common library used in machine learning projects is Tensorflow, an open-source deep learning library created by Google which can be used with both the Python and C++ programming languages. The library also has a framework called Tensorflow Lite, which allows for deployment on mobile devices, including both Android and IOS devices (Tensorflow Lite, 2019), which makes it a useful software to use for creating the pedestrian traffic light neural network.

A common problem faced in deploying a neural network on mobile devices is that the neural networks are unable to run quickly or efficiently enough on the device due to the low Central Processing Unit (CPU) power. When this occurs, the models are often "quantized," meaning that the precision of the numbers used in the neural network is decreased (Jacob, Klygis, Chen, Zhu, Tang, Howard, Kalenichenko, 2018). This slightly decreases the accuracy of the model, as the parameter values are not as exact; however, it can lead to large boosts in CPU usage and memory, greatly increasing the model's efficiency.

Conclusion

Although there have been many attempts to make pedestrian traffic lights accessible for the visually impaired, the current systems are still ineffective at fully ensuring the safety of the pedestrians. Thousands of accidents occur in which pedestrians are killed each year, and a large percentage of blind pedestrians have been involved in accidents (Al-Fuqaha, Oh, Kwigizile, 2018). Because of this, there is a clear need for a more effective solution to this problem, and the answer may lie in using machine learning techniques to ensure the safety of these visually impaired pedestrians.

A convolutional neural network that is able to accurately detect pedestrian traffic lights and detect their statuses could greatly assist visually impaired pedestrians, as they will receive direct input about whether or not it is safe to cross, rather than be forced to rely on audible signals, which often confuse the pedestrians (Kistler, 2015). The model would need to have high accuracy on the validation sets, meaning that it would need to be trained on a large dataset of images and tested on diverse images in order to ensure its accuracy with data that it has not previously been exposed to. A potential effective implementation of the neural network would be on cellular devices, as a large percentage of people carry these around wherever they go. This would allow for a convenient and easy to use method for visually impaired pedestrians to utilize the neural network.

References

Kistler, S. E. (2015, September 17). Why do intersections have different audible crosswalk signals? Retrieved from https://www.dispatch.com/article/20150911/NEWS/309119629

Pandey, P. (2019, March 28). Understanding the Mathematics behind Gradient Descent. Retrieved from https://towardsdatascience.com/understanding-the-mathematics-behind-gradient-descent-dde5dc9be06e

Parmar, R. (2018, September 02). Common Loss functions in machine learning. Retrieved from https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23

Team, C. (2016, June 29). Jaywalking defined and explained with examples. Retrieved from https://legaldictionary.net/jaywalking/

Sharma, A. (2017, March 30). Understanding Activation Functions in Neural Networks. Retrieved from https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0

Thrun, S. B. (1992). Efficient exploration in reinforcement learning. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.45.2894

Lapedes, A. S., & Farber, R. M. (1988). How neural nets work. In Neural information processing systems (pp. 442-456).

Retting, R. (2017). Pedestrian traffic fatalities by state. Governors Highway Safety Association, Washington, DC.

Pharr, J., Coughenhour, C., & Bungum, T. (2013). Environmental, human and socioeconomic characteristics of pedestrian injury and death in Las Vegas, NV. International Journal of Sciences.

Murphy, E. F. (1965). Some notes on canes and cane tips. Bull. Prosthet. Res, 10, 4.

Al-Fuqaha, A., Oh, J. S., & Kwigizile, V. (2018). 16-06 Vehicle-to-Device (V2D) Communications: Readiness of the Technology and Potential Applications for People with Disability.

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011, June). Flexible, high performance convolutional neural networks for image classification. In Twenty-Second International

Joint Conference on Artificial Intelligence.

Liu, W., Wen, Y., Yu, Z., & Yang, M. (2016, June). Large-margin softmax loss for convolutional neural networks. In ICML (Vol. 2, No. 3, p. 7).

Hassan, S. E., & Massof, R. W. (2012). Measurements of street-crossing decision-making in pedestrians with low vision. Accident; analysis and prevention, 49, 410–418. doi:10.1016/j.aap.2012.03.009

Moore, A. W. (2001). Cross-validation for detecting and preventing overfitting. School of Computer Science Carneigie Mellon University.

Technoaddict. (2018, August 08). 5 Most Used IDE For Developing Android Apps. Retrieved from https://medium.com/@technoaddict/5-most-used-ide-for-developing-android-apps-2aa601cb23ca

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., ... & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2704-2713).

TensorFlow Lite. (2019). Retrieved from https://www.tensorflow.org/lite

Vani, S. T. (2017, September 21). The pedestrian signal. Retrieved from https://medium.com/@sumukhatv/the-pedestrian-signal-11426f4c836d

Everything you need to know about Neural Networks. (2017). Retrieved from https://hackernoon.com/everything-you-need-to-know-about-neural-networks-8988c3ee4491

300mm Green Walk Man Pedestrian Crossing Light For Road Safety - Buy Pedestrian Crossing Light,Pedestrian Traffic Light,Led Traffic Light Product on Alibaba.com. (n.d.). Retrieved from https://www.alibaba.com/product-detail/300mm-green-walk-man-pedestrian-crossing_60756182483.html

New York 'Walk / Don't Walk' Sign. (n.d.). Retrieved from https://acesspeedshop.co.uk/product/new-york-walk-dont-walk-sign/

Szeto, A. Y., Valerio, N. C., & Novak, R. E. (1991). Audible pedestrian traffic signals: Part 3. Detectability. *Journal of rehabilitation research and development*, *28*(2), 71-78.