**Literature Review**

**Research Question: How are deep learning techniques implemented in autonomous vehicles to detect pedestrians?**

**Introduction**

Motor vehicles pose a serious threat to pedestrians, as collisions can often have serious repercussions. From 2011 to 2021 in the US there were 68,000 pedestrians killed, and 771,000 injured from motor vehicle accidents (Flint & Flint, 2023). Tefft (2013) shows that the possibility of severe injury or death increases drastically based on the vehicles speed. This means that a faster reaction time to slow down the speed of the collision could help to reduce the number of both fatalities and serious injuries. Singh (2018) found that the critical reason for about 92% of motor vehicle accidents were driver related. The most common category was due to recognition errors such as lack of attention from the driver or the driver being distracted. Choudhary & Velaga (2017) and Haque & Washington (2014) found that distractions from cell phones can increase the time of a drivers reaction to a pedestrian crossing the street by 40% for a simple conversation, up to 204% for complex texting. It’s expected that autonomous vehicles will reduce the amount of pedestrian collisions by removing the possibility of a distracted driver. This literature review focuses on the deep learning methods being used by autonomous vehicles to detect pedestrians, and provide insight on potential areas of study that can help to further enhance this technology.

**Deep Learning for Pedestrian Detection**

Deep learning plays an important role in autonomous vehicles for detecting pedestrians. An autonomous vehicle must do more than just see the location of a pedestrian to avoid collisions, but also understand where the pedestrian is moving, whether or not they will step in the vehicles path, how much time the vehicle has to stop, or the safest route to maneuver around the pedestrian if stopping is not an option. There are large amounts of data generated from the vehicle systems such as LiDAR, RGB images, or infrared images which can all provide useful information to the vehicle for this decision making process, and be used to supplement each other in the case of poor conditions such as rain, fog, or darkness (Vargas et. Al, 2021).

Convolutional neural networks are a powerful tool in object detection, and are being widely used for all computer vision tasks. Parameter sharing means they have reduced computation cost and network complexity, and they have shown great results in performing object detection tasks (Hakim & Fadhil, 2021). Two popular network types being used for pedestrian detection are regional proposal networks and You Only Look Once (YOLO) networks, as they are both capable of real-time efficiency in handling pedestrian detection tasks (Janai et. Al, 2021).

**Popular Pedestrian Detection Networks**

Regional proposal networks such as the Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et. Al, 2017) are popular networks for object detection. The network takes the image and proposes regions of interest based on the feature map extracted. The region pooling layer then removes overlapped regions, and converts them to a fixed size for the fully connected layers (Ibraheam et. Al, 2021). The last fully connected layer is a softmax to determine the probability of each target object existing within the proposed region. Li et. Al (2019) created an illumination aware faster R-CNN model for detecting pedestrians. They found that pedestrian detection had decreasing confidence with less illumination for both RGB and thermal imaging, and proposed a model to account for the amount of illumination when measuring the detection confidence level. This study highlighted the effectiveness of the faster R-CNN model for real time detection incorporating both RGB and thermal data to tackle the challenge of passenger detection without ideal lighting.

YOLO networks (Redmon et. Al, 2016) operate differently from regional proposal networks. Rather than breaking the image into different regions to analyze separately, YOLO uses a single neural network for predictions based on the full image. It solves object detection as a linear regression problem to spatially separated bounding boxes. The YOLO network has some advantages over the regional approach due to the fact it’s a single neural network, meaning that it can be trained to improve performance from end to end (Cheng, 2020). The objects are also not limited by the boundaries, giving less false positives in the background of an image. Nizar et. Al (2020) used the YOLO network to develop short range pedestrian detection with an RGB camera. The implemented system was able to detect 100% of pedestrians with 74.14% accuracy in position. It was successful in avoiding collisions with a human 85.71% of the time, with a 100% success rate at distances greater than 2 meters. Geng and Yin (2020) used the YOLO-V3 network to develop a human gesture detection system for autonomous vehicles by using infrared sensors. This allows for gesture detection in poor visibility such as fog, rain, or darkness which may limit capabilities of systems using RGB cameras.

**Current Challenges**

While autonomous vehicles provide the opportunity to reduce pedestrian collisions, there are still many challenges involved which make it difficult to develop and implement successfully. With pedestrian tracking it is important for the algorithm to run quickly for real time analysis, while also maintaining high accuracy to ensure confidence in the system (Yang et. al, 2022). There is usually a trade-off between accuracy and speed, presenting the challenge of how to process the data for detecting all objects (including pedestrians) without sacrificing accuracy for real-time algorithm speeds. Building trust in autonomous vehicles can be difficult to achieve (Morra et. al, 2019), and even a small amount of negative experiences with the technology can greatly impact the publics trust in the technology and slow down the rate at which it is adopted (Mesch & Dodel, 2022). Further studies on the real time performance while utilizing new and emerging technology can provide a better understanding of where we currently stand on this trade off.

Detecting a pedestrian that is partially occluded, or partially blocked from view of the sensors, is a challenging proposition for autonomous vehicles. There is little consistency in the definition of occlusion across different studies (Gilroy et. al, 2023), however there is a clear decrease in the accuracy of pedestrian detection models as the percentage of occlusion increases. The trade off of accuracy for faster speeds was also found to have a significant effect on partially occluded pedestrians caused high false-positive rates and lower detection confidence for occluded individuals (Gilroy et. al, 2023). Developing a consistent definition for levels of occlusion and focusing efforts on improved detection strategies in areas where a pedestrian could be blocked from view (such as driving in a lane next to parked vehicles) is crucial to improving pedestrian safety, as it will result in more consistency across studies in evaluating at what level of occlusion the system is no longer able to reliably detect a pedestrian.

Pedestrians have a tendency to be extremely unpredictable, and it can be difficult to determine their intent (Wang et. al, 2023). They often change their movement speed and direction, which provides a challenge in determining where they will move next or how quickly they will get there. Many challenging use cases don’t exist in current training sets (Coppola et. al, 2023). This limits the types of scenarios that a system can be prepared for, and makes it difficult to evaluate how a system will perform under difficult circumstances. There is also no universal simulator to test different modules in an autonomous vehicle all at once (Hussain & Zeadally, 2019). They are tested separately in different simulators, which can make it difficult to judge and test complicated requirements. Developing robust and reliable autonomous systems requires addressing these challenges to ensure the technology's effectiveness in diverse and unpredictable real-world environments.

In some cases when a collision may be unavoidable due to late detection of a pedestrian, the system may have to make a decision to protect the driver by colliding with the pedestrian, or risk the driver and swerve to miss the pedestrian. While avoiding this scenario with fast accurate data is the ideal situation, an autonomous vehicle will need to be trained on how to correctly make this decision based on all available sensor data (Valerdi, 2017). This type of moral dilemma is likely to have legal implications over the liability from the decision made by the system, and greatly impact the amount of trust the public has in autonomous vehicles. People will not want to use autonomous vehicles that risk the drivers life to protect a pedestrian that ran into the cars path, but the decision making system should prioritize saving both lives by missing the pedestrian when possible to do so. This is the type of extreme scenario that presents a major challenge to pedestrian detection in autonomous vehicles, as the ‘correct’ action may be perceived differently by different users.

**Focus For Future Work**

Despite the vast amounts of time and resources that have been invested into studying autonomous vehicles, there are more areas of pedestrian detection to be improved on in order to develop a reliable system. Many of the articles on pedestrian detection methods focus on the use of a single type of data input, such as an RGB camera. While some studies such as Geng and Yin (2020) combine data from multiple sensors to provide a pedestrian tracking system, there are a lack of studies that combine all available information (Such as LiDAR, RGB Cameras, Infrared, Radar, etc.) into a single system used for pedestrian detection. This is could be due to the lack of publicly available datasets containing all of the necessary information, but for implementation into a real-world scenario it is important to evaluate how the system performs with all available data.

Pedestrian tracking models were able to follow the location of a pedestrian, but did not analyze the likelihood of a pedestrian moving into the vehicles path at a future time. Bauer et. al (2023) used 3D human pose estimation on RGB and LiDAR images to determine a pedestrians pose, but little work has been done to apply the pose estimation results of a pedestrian to determine if they are moving into the vehicles path. Similarly, the presence of other objects such as parked cars, cross walks, or intersections should be considered as they increase the likelihood of a pedestrian stepping out onto the road. Comprehensive studies on how pedestrians behave in different environments or different situations could provide valuable insight for autonomous vehicles in predicting their behaviour. Continued research into these areas can help make fully autonomous vehicles safer for vulnerable road users such as pedestrians.

**References:**

Bauer, P., Bouazizi, A., Kressel, U., Flohr, F. (2023) "Weakly Supervised Multi-Modal 3D Human Body Pose Estimation for Autonomous Driving," Available From <https://ieeexplore-ieee-org.uniessexlib.idm.oclc.org/document/10186575>

Cheng, R.. (2020). A survey: Comparison between Convolutional Neural Network and YOLO in image identification. Journal of Physics: Conference Series, 1453 (1), 012139. Available from <https://doi.org/10.1088/1742-6596/1453/1/012139>.

Choudhary, P., Velaga, N. (2017). Modelling driver distraction effects due to mobile phone use on reaction time, Transportation Research Part C: Emerging Technologies, Available From: <https://doi.org/10.1016/j.trc.2017.02.007>.

Coppola, A., D'Aniello, C., Pariota, L., Bifulco, G. (2023). Assessing safety functionalities in the design and validation of driving automation, Transportation Research Part C: Emerging Technologies, Available From: <https://doi.org/10.1016/j.trc.2023.104243>.

Flint, R.D. and Flint, R.M.. (2023). Pedestrian fatalities in the United States: 2011–2021. Academic Emergency Medicine. Available from <https://doi.org/10.1111/acem.14766>.

Geng, K. and Yin, G.. (2020). Using Deep Learning in Infrared Images to Enable Human Gesture Recognition for Autonomous Vehicles. IEEE Access, 8, 88227–88240. Available from <https://doi.org/10.1109/access.2020.2990636>.

Gilroy, S., Mullins, D., Parsi, A., Jones, E., Glavin, M. (2023). Replacing the human driver: An objective benchmark for occluded pedestrian detection, Biomimetic Intelligence and Robotics, Available From <https://doi.org/10.1016/j.birob.2023.100115>.

Girshick, R.. (2015). Fast R-CNN. (2015). Available from <https://doi.org/10.1109/iccv.2015.169>.

Hakim, H. and Fadhil, A.. (2021). Survey: Convolution Neural networks in Object Detection. Journal of Physics: Conference Series, 1804 (1), 012095. Available from <https://doi.org/10.1088/1742-6596/1804/1/012095>.

Haque, M., Washington, S. (2014) A parametric duration model of the reaction times of drivers distracted by mobile phone conversations, Accident Analysis & Prevention, <https://doi.org/10.1016/j.aap.2013.09.010>.

Hussain, R. and Zeadally, S.. (2019). Autonomous Cars: Research Results, Issues, and Future Challenges. IEEE Communications Surveys & Tutorials, 21 (2), 1275–1313. Available from <https://doi.org/10.1109/comst.2018.2869360>.

Ibraheam, M., Li, K., Gebali, F., Sielecki, L. (2021). A Performance Comparison and Enhancement of Animal Species Detection in Images with Various R-CNN Models. AI, 2 (4), 552–577. Available from <https://doi.org/10.3390/ai2040034>.

Janai J, Güney F, Behl A, Geiger A. (2021) Computer vision for autonomous vehicles: Problems, datasets and state of the art. Available From <https://www.proquest.com/working-papers/computer-vision-autonomous-vehicles-problems/docview/2074461578/se-2>.

Li, C., Song, D., Tong, R., Tang, M. (2019). Illumination-aware faster R-CNN for robust multispectral pedestrian detection. Pattern Recognition, 85, 161–171. Available from <https://doi.org/10.1016/j.patcog.2018.08.005>.

Mesch, G., Dodel, M. (2023). The Acceptance of Driverless Cars: The Roles of Perceived Outcomes and Technology Usefulness. American Behavioral Scientist, Available From <https://doi-org.uniessexlib.idm.oclc.org/10.1177/00027642221127250>

Morra, L., Lamberti, F., Prattico, G., Rosa, S., Montuschi, P. (2019). Building Trust in Autonomous Vehicles: Role of Virtual Reality Driving Simulators in HMI Design. IEEE Transactions on Vehicular Technology, 68 (10), 9438–9450. Available from <https://doi.org/10.1109/tvt.2019.2933601>.

Nizar, T.N., Hartono, R. and Meidina, D.. (2020). Human Detection and Avoidance Control Systems of an Autonomous Vehicle. IOP Conference Series: Materials Science and Engineering, 879 (1), 012103. Available from <https://doi.org/10.1088/1757-899x/879/1/012103>.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016. Available from <https://doi.org/10.1109/cvpr.2016.91>.

Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (6), 1137–1149. Available from <https://doi.org/10.1109/tpami.2016.2577031>.

Singh, S. (2018). Critical Reasons for Crashes Investigated in National Motor Vehicle Crash Causation Survey, Crash Stats. Available From: <https://trid.trb.org/view/1507603>

Tefft, B. (2013). Impact speed and a pedestrian's risk of severe injury or death, Accident Analysis & Prevention, Volume 50, Pages 871-878, Available From: <https://doi.org/10.1016/j.aap.2012.07.022>.

Valerdi, R. (2017) The Next Challenge For Driverless Cars. Available From <https://web-p-ebscohost-com.uniessexlib.idm.oclc.org/ehost/pdfviewer/pdfviewer?vid=0&sid=94a9c875-b5e3-4bd7-a6de-4401f0400377%40redis>

Vargas, J., Alsweiss, S., Toker, O., Razdan, R., Santos, J. (2021). An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions. Sensors, 21 (16), 5397. Available from <https://doi.org/10.3390/s21165397>.

Wang, H., Jin, L., He, Y., Huo, Z., Wang, G., Sun, X. (2023). Detector–Tracker Integration Framework for Autonomous Vehicles Pedestrian Tracking. Remote Sensing, 15 (8), 2088. Available from <https://doi.org/10.3390/rs15082088>.

Yang, S., Chen, Z., Ma, X., Zong, X., Feng, Z. (2022). Real-time high-precision pedestrian tracking: a detection–tracking–correction strategy based on improved SSD and Cascade R-CNN. Journal of Real-Time Image Processing, 19 (2), 287–302. Available from <https://doi.org/10.1007/s11554-021-01183-y>.