

Verkeersveiligheidssystemen voor fietsers zijn ook mogelijk

Anoek Strumane – UGent IPI – anoek.strumane@ugent.be

Johan De Mol – UGent IPI – Johan.DeMol@UGent.be

Peter Veelaert - UGent IPI - peter.veelaert@ugent.be

Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 25 en 26 november 2021, Utrecht

Samenvatting

Voor het ontwikkelen van autonome voertuigen zijn vele nieuwe innovatieve technieken nodig. Deze technologieën focussen op het verhogen van de veiligheid en het rijgemak van weggebruikers; dit gebeurt uiteraard vanuit het standpunt van het voertuig. Opmerkelijk is dat ondanks de toenemende populariteit van fietsen in stedelijke gebieden er weinig onderzoek wordt gedaan om dergelijke systemen te ontwikkelen voor fietsers.

Het grootste obstakel bij de ontwikkeling van soortgelijke systemen voor fietsers is het voorzien van voldoende stroom nodig om de rekencapaciteit van de veiligheidssystemen efficiënt te laten werken. Op gemotoriseerde voertuigen levert de batterij voldoende stroom voor de goede werking van de CPU/GPU. Voor voertuigen is het trouwens mogelijk om duurdere veiligheidssystemen te voorzien.

In deze paper wordt de mogelijkheid onderzocht of voor de fietssystemen kunnen ontwikkeld worden die, door het voorspellen van mogelijke gevaren, de veiligheid van de fietser verhogen.

Het ontwikkelen van een systeem dat onveilige situaties voor de fietser voorspelt, kan voor de fietser een extra bescherming bieden. Dergelijke systemen kunnen een breed toepassingsgebied hebben. Zo kan de fietser gewaarschuwd worden voor een potentieel gevaar en kan het betreffende voertuig via signalen (licht/alarm) hierop attent gemaakt worden. (bv om dode hoekongevallen te voorkomen). Voor juridische doeleinden zou het ook gebruikt kunnen worden als een systeem dat de opname van bewijsmateriaal via camerabeelden triggert wanneer een gevaarlijke situatie voorspeld wordt.

1. Situering onderzoek

Binnen een Integraal beleid voor verkeersveiligheid worden klassiek maatregelen binnen het gebied van mens, voertuig en infrastructuur, ingedeeld in de drie E's: Engineering (in-frastructuur, voertuigtechnologie), Enforcement (handhaving) en Education (educatie, sensibilisering). In Vlaanderen worden er twee E's aan toegevoegd: Evaluatie¹ en Engagement (De Mol 2008).

Zeker bij voertuigtechnologie was dit voornamelijk op gemotoriseerd verkeer gericht. Bij voertuigtechnologie wordt ondermeer gedacht aan de inbouw van allerlei veiligheidverhogende technieken², bescherming inzittenden. Sommige van deze technieken zullen optimaal moeten werken om de zelfrijdende auto mogelijk te maken. Opmerkelijk is dat deze technieken niet worden voorzien/onderzocht voor andere weggebruikers. In casu voor de fiets zijn er geen veiligheidssystemen aanwezig die de fietser bv. attent maken op mogelijk gevaar of het rijgedrag laten aanpassen.

In deze paper wordt de mogelijkheid onderzocht of er veiligheidssystemen voor de fietser kunnen ontwikkeld worden waardoor de kans op een ongeval sterk kan worden verminderd. Daarbij zal een model moeten ontwikkeld worden dat het gevaar voor een ongeval kan voorspellen.

2. Inleiding

In deze paper wordt onderzocht of een systeem ontwikkeld kan worden dat gevaarlijke situaties voorspelt voor fietsers. De gevaarlijke situaties die voorspeld worden zijn scenario's waarbij de kans op een botsing tussen een groot gemotoriseerd voertuig en de fietser hoog is. Een dergelijk systeem zou ongevallen kunnen voorkomen door de fietser of het de bestuurder van het voertuig tijdig te waarschuwen, zodat deze een gepaste actie kan ondernemen. Het systeem zou ook voor juridische doeleinden kunnen worden gebruikt door het te verbinden aan een *dash cam* systeem. Daarnaast is het ook toepasbaar om de veiligheid van verschillende verkeerslocaties in kaart te brengen, om zo het tekort aan informatie over de veiligheid van fietsinfrastructuren aan te pakken.

Er wordt een eenduidige definitie gegeven aan een gevaarlijke situatie afhankelijk van de snelheid en nabijheid van voertuigen. Aan de hand van deze definitie worden verschillende scenario's gelabeld en deze aanduiding van al dan niet gevaarlijk wordt zo ver mogelijk in de tijd voorspeld. Er wordt in dit onderzoek gebruikgemaakt van LiDAR of *Light Detection And Ranging* sensoren als enige invoer van data. Deze LiDAR scans worden omgezet naar polaire *occupancy grids* die in paren van twee aan een neurale netwerk worden gevoed.

¹ https://assets.vlaanderen.be/image/upload/v1626161661/20210709_MOW_Verkeersveiligheidsplan_Vlaanderen_2021-2025_DEF_nrwbw7.pdf

² Voorbeelden: Collision Avoidance/Warning, Intelligent Speed Adaptation, Alcolock Key, Blind Spot Detection, Lane Changing Support, Predictive Brake Assist, Drowsy Driver Detection, Lane Keeping Assistant, Night Vision, ACC (Advanced Cruise Control), AEB (Automatic Emergency Braking), ISA (Intelligent Speed Adaptation).

Er worden veelbelovende resultaten bereikt. Zo kunnen meer dan de helft van de gevaarlijke situaties tijdig en correct worden opgespoord met een lage verhouding aan valse alarmen. Deze paper is gebaseerd op een onderzoek dat in het kader van een masterthesis³ werd uitgevoerd.

2.1 Doel

Als eerste stap in het proces moet er gedefinieerd worden wat gevaar is. Dit moet op een gepaste manier gebeuren zodanig dat de gebruikte data op een eenduidige manier gelabeld kan worden als al dan niet gevaarlijk. Deze aanduiding van gevaar zal dan voorspeld worden door een neurale netwerk, zodat de fietser gewaarschuwd wordt.

Opdat een fietser zou kunnen reageren op een waarschuwing van gevaar, moet de fietser snel een signaal ontvangen zodat hij/zij tijdig ingelicht is over een naderende gevaarlijke situatie. In (Buchholtz, 2020) wordt er geconstateerd dat de gemiddelde reactiesnelheid van een fietser ongeveer 0,4 seconden bedraagt met 0,7 seconden als maximaal waargenomen tijd. Om genoeg marge te hebben voor het verwerken van de invoergegevens, wordt er getracht om de veiligheidssituatie van minstens één seconde verder te voorspellen. Op deze manier is er zeker 0.7 seconden tijd tussen een waarschuwing van het systeem en het effectief optreden van gevaar.

Ten slotte moet het systeem dat ontwikkeld wordt zo weinig mogelijk rekenkracht en elektrisch vermogen vereisen. Zo focussen bijna alle onderzoeken in verband met het voorspellen en voorkomen van gevaarlijke verkeerssituaties zich op een zo correct mogelijke voorspelling, zonder veel rekening te houden met de rekenkracht die het systeem zal nodig hebben. Dit omdat men ervan uitgaat dat het systeem op een gemotoriseerd voertuig gemonteerd wordt, waardoor het toegang heeft tot voldoende energie en het haalbaar is om krachtige GPU's of CPU's te gebruiken. Enkele voorbeelden van dergelijke systemen worden voorgesteld in (Luo, 2018), (Wang, 2017) and (Djuric, 2020). Dit soort systemen zijn echter onhaalbaar bij fietsen. Zo zou het weinig interessant zijn om een relatief groot en duur systeem op een fiets te zetten die zelf maar enkele honderden euro's kost. Bij het ontwerp keuzes ligt de focus daarom, naast het zo goed mogelijk voorspellen van de gevaarlijke situaties, voornamelijk op het zo licht mogelijk houden van het systeem.

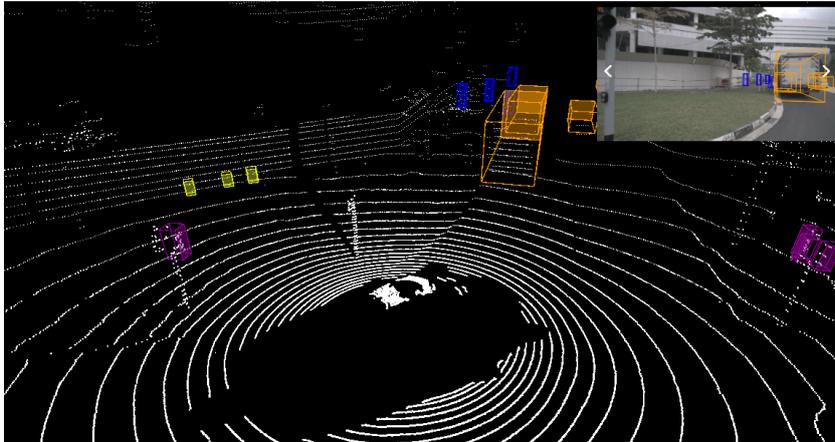
3. Terminologie

3.1 LiDAR

Light Detection And Ranging of LiDAR is een techniek om omgevingen in kaart te brengen die zeer gelijkaardig is aan RADAR. Het grootste verschil tussen RADAR en LiDAR is dat LiDAR gebruikmaakt van lichtpulsen in de plaats van radiogolven. Deze lichtpulsen worden de omgeving ingestuurd en vervolgens wordt gemeten hoelang het duurde voor elke puls om terug bij de LiDAR sensor te geraken. Op deze manier kunnen de coördinaten van alle punten in de omgeving berekend worden en kan er een puntenwolk gemaakt worden zoals te zien in Figuur 1. Elk punt in deze wolk bestaat uit

³ Voor het behalen van een Master in de industriële wetenschappen: informatica.

een driedimensionaal coördinaat en een intensiteitswaarde dat aangeeft met welke intensiteit de lichtpuls terugkeerde.



Figuur 1: voorbeeld LiDAR point cloud met annotaties van omliggende objecten

3.2 Artificiële neurale netwerken

Artificiële neurale netwerken zijn binnen de computerwetenschappen een van de manieren om een computer te laten *leren*. Om te leren heeft een neuraal netwerk een grote dataset nodig bestaande uit de invoergegevens en hun bijhorende labels. Stel dat het doel van het netwerk is om foto's van honden en katten te classificeren, dan zijn de invoergegevens de te classificeren foto's en de labels zijn 'kat' of 'hond'. Na het zien van enorm veel voorbeelden van foto's van katten en honden en hun respectievelijke labels, zal het neuraal netwerk leren veralgemenen en bepaalde zaken herkennen, zoals de vorm van de oren van het dier.

Neurale netwerken zijn complexe structuren met een groot aantal parameters die veel verschillende waarden kunnen aannemen. Het bouwen van dergelijke netwerken is dus niet vanzelfsprekend. In dit paper gaan we net in op de werking en architectuur van neurale netwerken, enkel op de resultaten. De belangrijkste elementen waar we rekening moeten mee houden is dat de complexiteit, rekenkracht en benodigde energie van een neuraal netwerk sterk stijgen naarmate het aantal invoergegevens groter wordt.

Belangrijk om weten, is dat in neurale netwerk stijgen benodigde energie en rekenkracht enorm sterk stijgt naarmate de invoergegevens groter zijn. Een afbeelding van bijvoorbeeld 10 op 10 pixels classificeren zal veel minder zwaar zijn dan een afbeelding van 256 op 256 pixels. Het is dus belangrijk om deze invoergegevens zo beperkt mogelijk houden.

4. Definitie van gevaar

4.1 De nood aan een definitie

Om een realistisch gevarensysteem te ontwikkelen is de beste optie het gebruik van real-time data van ongevallen en gevaarlijke situaties. Deze soort data is echter zo goed als onbestaand als het gaat over fietsers. Binnen het tijdsbestek van enkele maanden is het

ook onhaalbaar zelf voldoende gevaarlijke situaties op te nemen en aan te duiden. Daarom wordt er gebruik gemaakt van een bestaande publieke dataset van verkeersomstandigheden. Er wordt een definitie gekozen waarmee deze invoerdata op een ondubbelzinnige manier kan worden gelabeld. Met een strikte definitie van gevaar kan het proces van labelen op een automatische manier gebeuren.

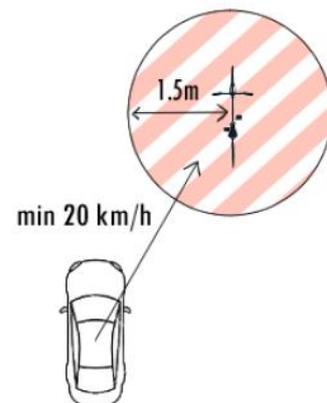
4.2 De definitie

Na een extensieve studie naar het voorkomen van fietsongevallen werd er besloten om gevaar te definiëren als een gemotoriseerd voertuig dat aan een minimumsnelheid richting de fietser beweegt en binnen een bepaalde afstand rond de fietser komt.

De minimumsnelheid van het gemotoriseerd voertuig dat mogelijk een gevaar kan vormen wordt op 20km/u ten opzichte van de fietser gezet. Dit omdat er zelden ongevallen gebeuren die leiden tot verwondingen onder een snelheid van 30km/u (Kim, 2007). Als men dan rekening houdt met het feit dat de fietser ook kan bewegen in dezelfde richting als de wagen, wordt zijn snelheid gecompenseerd door van deze 30km/u een marge van 10km/u af te trekken.

In België moet er volgens het verkeersreglement ten allen tijden minstens 1,5m afstand gehouden worden tussen wagens en fietsers. Binnen de bebouwde kom wordt deze afstand verlaagd naar 1m.⁴ Aangezien deze afstanden met zorg zijn gekozen, zijn deze al een goede richtlijn om als straal te gebruiken om een gevaarlijke situatie te definiëren. Deze afstanden zijn helaas te klein om als criterium te gebruiken wanneer de invoerdata afkomstig is van een LiDAR sensor gemonteerd op een personenwagen. Daarom wordt de gebruikte afstand voor de doeleinden van deze paper verhoogd naar 4m. Deze afstand werd zo gekozen dat deze zo klein mogelijk bleef terwijl er nog steeds genoeg *onveilige* verkeerssituaties gevonden worden.

Personenwagens die aan deze criteria van minimumsnelheid en afstand voldoen, worden beschouwd als gevaarlijk. Hierbij wordt opgemerkt dat dit niet enkel voertuigen zijn die recht op de fietser afkomen, maar ook voertuigen die de fietser bijvoorbeeld zijdelings zullen kruisen, zolang deze aan een *relatieve* snelheid van 20km/u de fietser benaderen. Een voorbeeld wordt gegeven in Figuur 2.



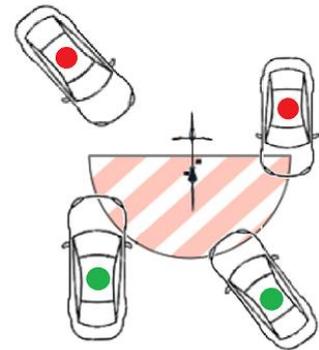
Figuur 2: voorbeeld van voertuig dat als gevaarlijk zal gezien worden als het een minimumstraal betreedt.

⁴ Idem dito in Frankrijk, in Duitsland (ausreichender Seitenabstand' aanhouden - §5, Absatz 4 van de Straßenverkehrsordnung.= minstens anderhalve meter), Engeland ("plenty of room = alsof fietser de breedte van een auto heeft), Spanje: 2 meter afstand en verlagen van de snelheid met 20 km/uur, ...

4.3 Aanduiding van gevaar

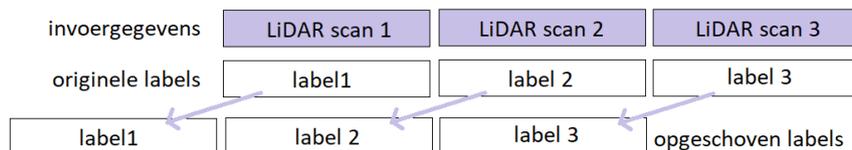
Om de data te labelen werd er gekozen om gebruik te maken van een binaire labeling. Dit betekent dat een situatie positief of negatief kan zijn, wat respectievelijk gevaarlijk of veilig aanduidt.

Een bijkomend probleem bij data verzamelen vanop een fiets, is dat een sensor nooit een zicht van 360° rond de fietser kan behalen. Dit kan enkel indien de sensor op de helm van de fietser is gemonteerd, maar dan staat deze echter niet stabiel genoeg om eenduidige data te verzamelen. Er wordt daarom gekozen om enkel rekening te houden met de achterste 180° van de fietser. Naderende voertuigen worden dus enkel als gevaarlijk gezien wanneer ze de fietser langs de achterste helft benaderen, zoals te zien in Figuur 3.



Figuur 3: voorbeelden van voertuigen die zullen gedetecteerd worden (groen) en voertuigen die niet zullen gedetecteerd worden (rood).

Deze aanduiding van gevaar moet ten slotte ook in de tijd worden opgeschoven, zodanig dat het systeem gevaar kan *voorspellen* in de plaats van gewoon *detecteren*. Als de LiDAR scan op tijdstip t het gevarenlabel van tijdstip $t+0,5s$ krijgt, dan zal het al dan niet optreden van gevaar binnen 0,5 seconden voorspeld worden. Er kan zo eenvoudig ook verder in de tijd voorspeld worden door de toegevoegde factor van 0,5s te verhogen. De modellen werden eerst getest met voorspelling van gevaar binnen 0,5 seconden, dan werd dit tijdsinterval stapsgewijs verhoogd tot 2,5 seconden om zo te bepalen tot hoever in de toekomst het systeem nog een interessante performantie kon halen.



Figuur 4: voorbeeld opschuiving labels. Bovenaan de invoergegevens met hun originele labels. Onderaan de opgeschoven labels.

5. Dataverwerking

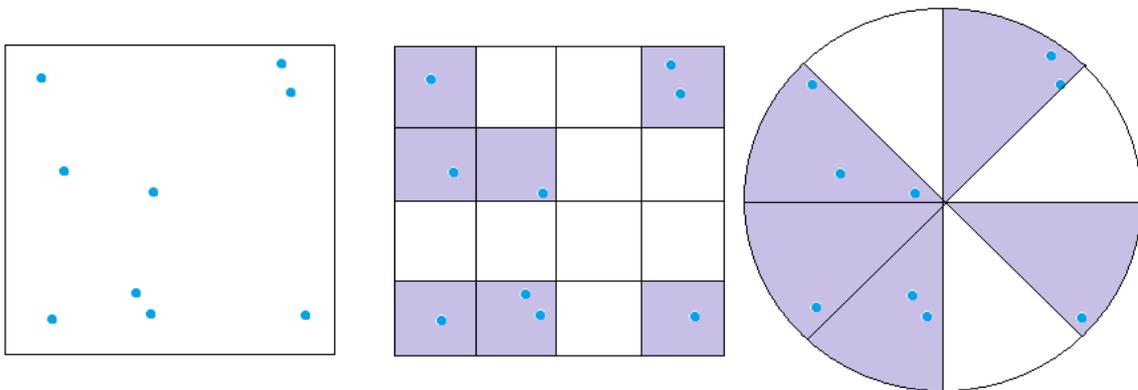
5.1 Origine van de data

Doordat LiDAR gegevens die opgenomen zijn vanop fietsen zo goed als onbestaand zijn, werd er gekozen om gebruik te maken van LiDAR data afkomstig van een personenwagen. Voor het trainen en testen van neurale netwerken maakt het niet veel uit of de data vanop een fiets of vanop een wagen werd opgenomen. Het belangrijkste is het gedrag van beide ten opzichte van elkaar in kaart brengen. De wagen waarop de sensor gemonteerd staat zal anders manoeuvreren dan een fiets, maar het voorspellen of een ander voertuig een al dan niet gevaarlijke manier de sensor zal naderen, blijft hetzelfde. Er zijn verscheidene gegevensbanken online beschikbaar die dit soort data bevatten. Er werd gekozen voor de NuScenes (Caesar, 2019) dataset, die een enorm

uitgebreide verzameling aan LiDAR scans bevat, in combinatie met overeenkomstige camera en RADAR beelden. Deze opnames zijn zeer goed geannoteerd. Daarbij bevat deze dataset een groot spectrum aan situaties, zo zijn er scenes in drukke steden, in rustige buurten, in volle zon, tijdens sneeuwval,...

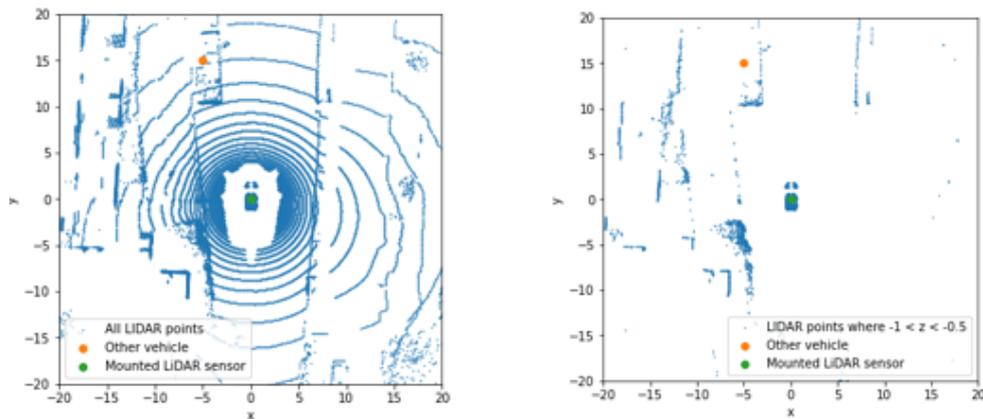
5.2 Rooster-representatie

Eén enkele LiDAR scan van de NuScenes gegevensbank bevat ongeveer 34.000 punten. Elk punt bestaat uit een 3-dimensionaal coördinaat dat de locatie van het punt beschrijft en een bijkomend nummer dat aanduidt met welke intensiteit de straal van het overeenkomstige punt terugkeerde. Om één ruwe scan te laten verwerken, zouden er dus $34.000 * 4 = 136.000$ getallen gevoed moeten worden aan een neurale netwerk. Een dergelijke LiDAR scan kan echter samengevat worden in een rooster formaat, waarbij samenliggende punten gegroepeerd worden in één cel. Cellen die geen punten bevatten, worden als *leeg* aangeduid en cellen die wel punten bevatten worden als *bezet* aangeduid. Dit zal ervoor zorgen dat het systeem minder geheugen verbruikt, minder zwaar is en een sterke snelheidsboost krijgt **Fout! Verwijzingsbron niet gevonden.**



Figuur 5: voorbeeld omzetting van punten naar een rooster formaat. Gekleurde cellen worden aangeduid als bezet. Links de originele punten. In het midden het cartesisch roosterformaat. Rechts het polair roosterformaat.

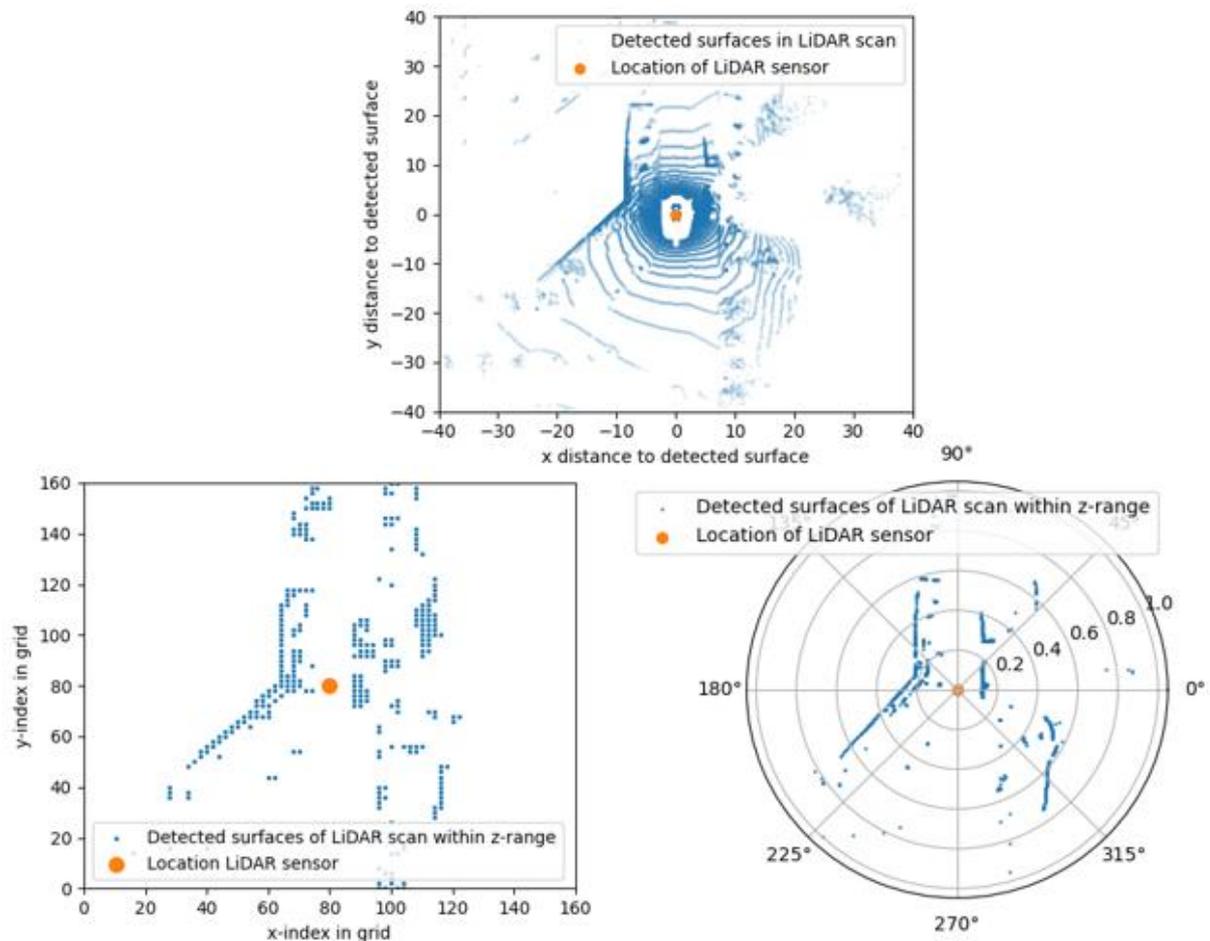
Een dergelijk rooster kan uit twee of drie dimensies bestaan. Bij het gebruik van drie dimensies gaat er weinig informatie verloren. Bij een tweedimensionaal rooster wordt de hoogte parameter, de z-as, achterwege gelaten. Er wordt dan enkel rekening gehouden met punten die binnen een bepaald hoogte-interval liggen en deze worden dan samengevat alsof ze op eenzelfde hoogte liggen. In Figuur 6 wordt het bovenaanzicht gegeven van alle LiDAR punten in vergelijking met de LiDAR punten binnen een bepaald hoogte-interval. Duidelijk is dat wanneer dit interval goed wordt gekozen, andere voertuigen nog duidelijk zichtbaar zijn en veel redundante informatie verwijderd wordt. Op deze manier verdwijnt de nood aan een z-parameter. Er werd gekozen om te werken met een tweedimensionaal rooster, omdat dit veel geheugen en rekenkracht bespaart. Daarbij komt dit soort rooster ook beter overeen met de goedkopere *single-beam* LiDAR sensoren op de markt.



Figuur 6: links de originele point cloud, rechts de punten van deze point cloud die op bepaalde hoogtes liggen.

Tenslotte kan een tweedimensionaal rooster op twee manieren worden voorgesteld, een cartesisch rooster of een polair rooster. Bij het gebruik van cartesische coördinaten, wordt elk punt, of in dit geval elke cel, voorgesteld door een (x, y) coördinaat. Bij een polaire voorstelling worden deze gegeven door een afstand en een hoek, (r, θ) . De twee opties worden weergegeven in Figuur 7. Bij een polair rooster is het redelijk eenvoudig om in elke cel enkel het dichtstbijzijnde punt bij te houden, waardoor punten die erachter liggen verdwijnen. Dit zorgt nogmaals voor een verkleining van het rooster, omdat er per θ maar één afstand bijgehouden wordt. Het weglaten van punten die verder liggen dan dit dichtste punt komt daarbij ook beter overeen met de LiDAR representatie van de besproken *single-beam* sensoren. Ook komt de quantisatie bij een polair rooster beter overeen met de quantisatie van de LiDAR sensor. Zo zullen is de resolutie lager op langere afstanden dan op korte afstanden. In een cartesisch rooster is de resolutie overal gelijk. Om deze redenen werd er gekozen om te werken met een polair rooster.

Door een hoekresolutie van 0,25 graden te nemen en enkel te kijken naar de achterste 180 graden, komt één rooster overeen met 720 elementen. Deze elementen bevatten elk één cijfer, namelijk de afstand tot het dichtste punt. Zo wordt de grootte van de originele LiDAR scan met een factor van ongeveer 200 gereduceerd.



Figuur 7: bovenaan de originele point cloud. Links het cartesisch rooster. Rechts het polair rooster.

5.3 Dynamische informatie

Wanneer er maar één enkele omgezette scan gebruikt wordt als input voor een neurale netwerk, kan er geen dynamische informatie uit deze invoergegevens gehaald worden. Daarom moest er een manier gekozen worden om opeenvolgende scans te combineren. De effectieve snelheid van objecten in de omgeving worden niet expliciet bepaald aangezien dit soort proces veel meer rekenkracht zou vereisen dan nodig en nuttig is. In de plaats daarvan worden opeenvolgende scans gecombineerd wanneer ze aan het neurale netwerk gevoed worden, zodat het netwerk zelf impliciet kan rekening houden met dynamische gegevens van de omgeving.

Er werden verschillende opties overwogen, namelijk: Dynamic Occupancy Grid Mapping of DOGMa (Schreiber, 2020), Recurrent Neural Networks of RNNs (Bianchi, 2017) en het aan elkaar plakken van opeenvolgende grids. DOGMas zijn een manier om aan de hand van neurale netwerken dynamische informatie te extraheren uit opeenvolgende roosters. RNNs zijn een subtype van neurale netwerken die rekening kunnen houden met de volgorde waarin input gegevens worden gevoed. Bij het zien van een nieuwe scan, kent de RNN nog informatie over de vorige scan(s) en kan hier dus ook nog rekening mee houden. Alhoewel ze toelaten om informatie te gebruiken die ver teruggaat in de tijd, zijn

DOGMA's en RNNs zeer complex, zijn ze moeilijker te trainen en vragen ze veel rekenkracht en geheugenruimte. Daarom werd gekozen voor een veel eenvoudiger netwerk waarbij we de data van opeenvolgende scans concateneren en als één geheel invoeren aan het neurale netwerk. Door twee scans samen te voegen tot één inputvector, verdubbelt de grootte van de roosters zodat elk rooster $720 \times 2 = 1440$ getallen groot is.

5.4 Labelen van roosters

Opdat een neurale netwerk zou kunnen leren, moeten het een groot aantal voorbeelden voorgeschoteld krijgen samen met de respectievelijke uitkomsten die ervan verwacht worden. Elk rooster dat als invoer gebruikt wordt, moet dus eerst een label krijgen dat aangeeft of de situatie aangeduid op dat moment al dan niet gevaarlijk is.

Bijgevolg zijn de annotaties van de NuScenes dataset in deze stap het belangrijkste. In de NuScenes databank heeft elke individuele scan, naast de LiDAR gegevens en de camerabeelden, ook annotaties voor alle voertuigen die in de scan aanwezig zijn. Elk voertuig heeft een token dat uniek is voor dat voertuig en dat hetzelfde blijft doorheen opeenvolgende scans. Zo kan een specifiek voertuig gevolgd worden tijdens het overlopen van de scans. Elk van de tokens heeft ook per scan een *bounding box*, namelijk de hoogte, breedte en lengte van het voertuig, een coördinaat van het centrum van het voertuig en een aanduiding van de oriëntatie van het voertuig. De snelheid van een specifiek voertuig ten opzichte van de LiDAR sensor op moment t kan afgeleid worden door zijn coördinaten over de voorbije scans $t-1$, $t-2$,... te combineren. De afstand tot de LiDAR sensor kan afgeleid worden door de bounding box te verschuiven naar de locatie van het centrum van de wagen en deze dan te roteren zodat deze overeenkomt met de oriëntatie van het voertuig. Daarna wordt de kortste afstand tot deze bounding box berekend zodat de afstand tot het voertuig gekend is.

Er werd al aangehaald dat het netwerk enkel rekening houdt met de achterste 180° van de sensor. Dit legt bijkomende restricties op voor het herkennen van gevaarlijke voertuigen. Oorspronkelijk waren de enige twee vereisten van een gevaarlijk voertuig dat het aan een relatieve snelheid van minimum 20km/u de sensor naderde en dat het voertuig dichter dan 4 meter bij de sensor kwam. Bijkomend moet het voertuig nu ook zichtbaar zijn in de achterste 180° de scans die aan t voorafgaan. Het voertuig moet ook volledig zichtbaar zijn. Dit betekent dat voertuigen die loodrecht vanaf de zijkant van de sensor de fietser naderen niet herkend zullen worden.

Met deze regels en de annotaties werd er vervolgens voor elk gemotoriseerd voertuig in elke scan eenduidig bepaald of het voertuig al dan niet gevaarlijk is. Deze labels worden dan aan toegekend aan de scans. Zo kan het bijhorend label van elk rooster worden opgehaald wanneer het trainingsproces gestart wordt. Deze labels worden dan terug in de tijd opgeschoven. Aangezien de LiDAR scans van NuScenes opgenomen worden met een frequentie van 2Hz, komt een opschuiving van 1 stap overeen met 0,5s. Dit wordt gedaan door het label van rooster r_t op te schuiven naar het rooster r_{t-1} waardoor het neurale netwerk zal leren om het eigenlijke label van rooster r_t te voorspellen wanneer het rooster r_{t-1} te zien krijgt.

5.5 Vergroting van de dataset

De volledige NuScenes dataset bevat 850 *scenes* die elk 40 *samples* bevatten. Eén sample is een momentopname van de omgeving en komt dus overeen met één LiDAR scan en bijhorende camera en RADAR beelden. Een scene is een opeenvolging van samples, genomen over 20 seconden tijd. Dit komt neer op een totaal van 34.000 samples of LiDAR scans. Wanneer de hierboven besproken manier van labeling hier echter op toegepast wordt, blijkt het dat er maar 122 gevaarlijke scans zijn. Dit is onvoldoende om een neuraal netwerk te trainen op gevaarlijke situaties. Daarom werd er *data augmentation* toegepast, wat neerkomt op het vergroten van de dataset door het creëren van fictieve gevaarlijke situaties. De augmentatietechnieken gebruikt in deze paper zijn grotendeels gebaseerd op de resultaten van (Hahner, 2020).

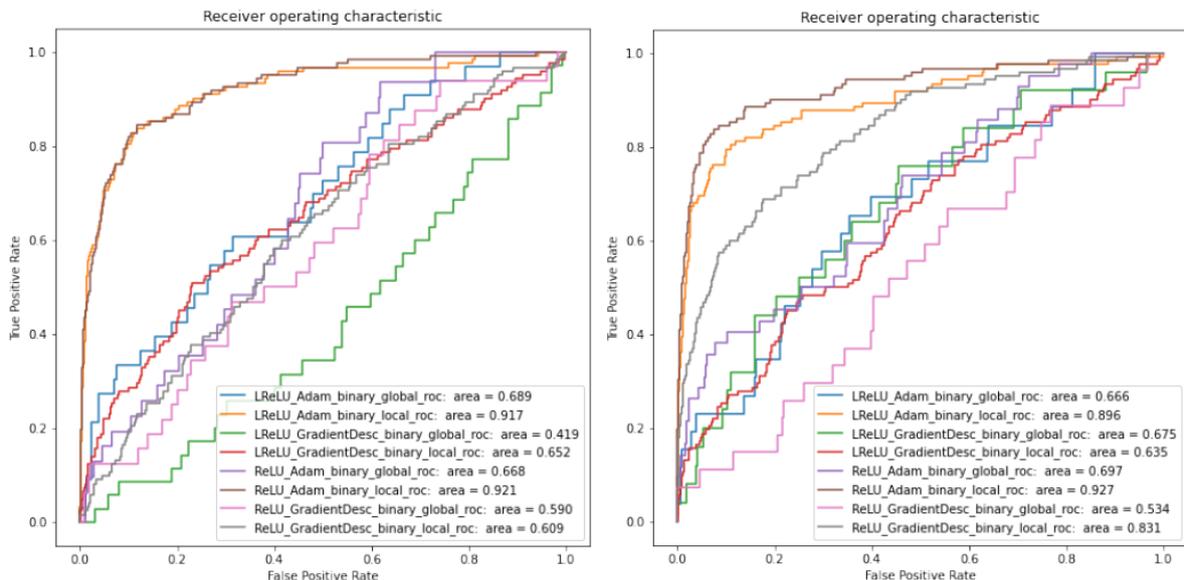
De dataset werd op twee manieren vergroot: local augmentation en global augmentation. In het globale geval, werd elke scan geroteerd, zodanig dat er telkens een andere fractie van het totaalzicht werd opgenomen in de zichtbare 180°. Deze rotaties gebeurde in stappen van 15° en gingen van minimaal -45° tot maximaal 45°. Grotere rotaties zouden leiden tot zeer onrealistische situaties, zoals de sensor dat loodrecht door het verkeer rijdt of zelfs in de tegengestelde richting. De scans werden ook gespiegeld om de as in de rijrichting van de sensor. Deze gespiegelde scans werden op hun beurt ook globaal geroteerd. Dit resulteerde in een totaal van 4880 gevaarlijke situaties. Bij de lokale variant werden de LiDAR punten van gevaarlijke voertuigen uit scans geknipt en in veilige scans geplakt, zodat deze dan als onveilig zou gezien worden. Hiervoor werden alle scans opnieuw geroteerd, dit keer in stappen van 30°, van 0° tot 360°. Uit deze geroteerde scans werden de LiDAR punten van gevaarlijke voertuigen verzameld en opgeslagen. Dit resulteerde in een totaal van bijna 30.000 gevaarlijke instanties, die in veilige scans konden geplakt worden om zo onveilige scans te creëren. De twee gebruikte manieren werden apart gebruikt bij het trainen van verscheidene netwerken om het verschil in performantie tussen de twee methodes weer te geven.

6. Modellen en resultaten

De gelabelde en met augmentatietechnieken uitgebreide dataset werden gebruikt om verschillende types van netwerken te trainen onder verschillende modaliteiten. Verschillende neurale netwerken werden getest, elk met zijn eigen unieke parameters. Twee belangrijke parameters hadden te maken met de gebruikte invoer: hoever in de toekomst voorspeld werd en of er gebruik werd gemaakt van lokale of globale augmentatie. Er werd ook geëxperimenteerd met verschillende types van netwerk optimizers en activatiefuncties. Twee verschillende netwerkarchitecturen werden getest: een eenvoudig *feed forward* neuraal netwerk of FFNN en een *convolutional* neuraal netwerk of CNN. Het grootste voordeel van CNNs is dat ze beter zijn in het *begrijpen* van ruimtelijke invoer, zoals de roosters gemaakt van LiDAR punten. Ze vereisen hierdoor echter ook meer rekenkracht en geheugen. Simpele FFNNs zijn de lichtste vorm van neurale netwerken, maar hebben op hun beurt meer moeite met ruimtelijke gegevens. Als optimizers werden *gradient descent* en de *adam optimizer* gebruikt. Als activatiefuncties werden de *ReLU* en de *Leaky ReLU* uitgeprobeerd.

In Figuur 8 en Figuur 9 worden respectievelijk de ROC-resultaten voor het CNN en het FFNN netwerk getoond wanneer er 0,5 seconden in de toekomst voorspeld werd. De volledige betekenis van de curve wordt achterwegen gelaten. Het belangrijkste bij deze curves is de oppervlakte onder de curve. Hoe groter deze oppervlakte, hoe correcter de voorspellingen van het neurale netwerk. Het is duidelijk dat in beide types neurale netwerken, deze met lokale augmentatie en de adam optimizer het beste werkten. De numerieke resultaten wanneer een test dataset werd gegeven aan de beste FFNN en CNN netwerken worden gegeven in Tabel 1. Hierbij moeten de kolommen van de tabel verder uitgelegd worden:

- *tp* of *true positives*: aantal roosters dat als gevaarlijk waren gelabeld die ook als gevaarlijk werden voorspeld door het netwerk
- *fp* of *false positives*: aantal roosters dat als veilig waren gelabeld die als gevaarlijk werden voorspeld door het netwerk
- *tn* of *true negatives*: aantal roosters dat als veilig waren gelabeld die ook als veilig werden voorspeld door het netwerk
- *fn* of *false negatives*: aantal roosters dat als gevaarlijk waren gelabeld die als veilig werden voorspeld door het netwerk



Figuur 8: ROC resultaten van de CNNs op 0,5s in de toekomst
 Figuur 9: ROC resultaten van de FFNNs op 0,5s in de toekomst

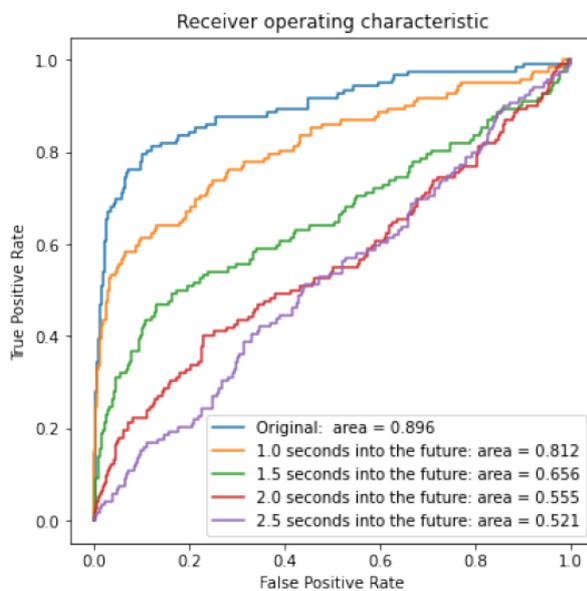
	tp	fp	tn	fn
FFNN	93	179	1350	29
CNN	86	90	1339	36

Tabel 1: numerieke resultaten van de beste CNN en FFNN architecturen.

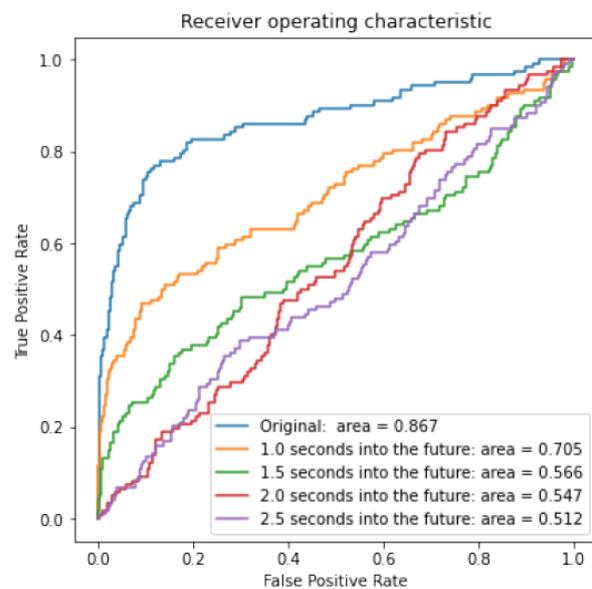
Er moet opgemerkt worden dat deze gegevens verkregen zijn op een heel ongebalanceerde testset. Het aantal veilige situaties ligt namelijk veel hoger dan het aantal gevaarlijke, wat beter aansluit bij de realiteit. Er is te zien dat het FFNN 93 van de 122 gevaarlijke situaties correct voorspelt, maar ook 179 roosters aanduidt als gevaarlijk wanneer ze dit niet zijn, zogenaamde valse alarmen. De CNN variant voorspelt iets minder gevaarlijke situaties correct, maar zijn verhouding ten opzichte van valse alarmen ligt veel beter. Deze valse alarmen zijn grotendeels te wijten aan de harde grenzen van

20km/u en 4 meter die werden opgelegd bij het definiëren van gevaar. Hierdoor wordt een voertuig dat net niet de limiet van 20km/u of 4m haalt, niet als gevaarlijk aangeduid door de labeling, maar kan het neurale netwerk dit wel als gevaarlijk zien.

De volgende stap was het verhogen van de tijd tussen de opgenomen scan en het waargenomen gevaar, zodanig dat het netwerk verder in de toekomst voorspelt. Er werd in stappen van 0,5s gewerkt van 0,5s tot 2,5 seconden. In Figuur 10 en Figuur 11 worden respectievelijk de ROC resultaten van het CNN en het FFNN getoond. Het is duidelijk dat de CNN een minder sterke daling van accuraatheid ziet dan het FFNN wanneer er verder in de toekomst wordt opgeschoven. De numerieke resultaten van het CNN worden in Tabel 2 weergegeven.



Figuur 11: ROC resultaten van de CNN wanneer er verder in de tijd wordt voorspeld.



Figuur 10: ROC resultaten van de FFNN wanneer er verder in de tijd wordt voorspeld.

seconden	tp	fp	tn	fn
0,5	86	90	1339	36
1,0	67	72	1432	55
1,5	53	162	1296	69
2,0	50	345	989	72
2,5	48	572	930	74

Tabel 2: numerieke resultaten van het best CNN wanneer er verder in de tijd wordt voorspeld.

Tot en met 1 seconde in de toekomst worden meer dan de helft van de gevaarlijke situaties correct voorspeld en ligt de verhouding $\$tp/fn\$$ nog op een aantrekkelijk punt. Wanneer het netwerk verder in de toekomst voorspelt kan het maar ongeveer $\$40\%\$$ van de gevaarlijke situaties correct voorspellen en daalt de verhouding $\$tp/fp\$$ drastisch. Deze sterke daling in accuraatheid is grotendeels te wijten aan het feit dat sommige wagens uit het zicht van de sensor verdwenen omdat deze op het moment van de LiDAR opname bijvoorbeeld nog achter een andere wagen verschild zaten. Daarbij wordt het correct voorspellen van het traject dat een auto zal nemen snel moeilijker naarmate het

voertuig verder weg is. Deze trajectvoorspelling complicaties vormen nog steeds problemen, zelfs bij geavanceerde systemen die enorme rekenkracht vereisen.

Ten slotte wordt er nog aangehaald dat het voorspellen van het al dan niet optreden van gevaar maar 0,6 milliseconden duurde op een laptop met een Intel i5 CPU (€170) en 8GB RAM-geheugen. Door een tekort aan testen is het helaas niet met zekerheid gegarandeerd, maar er is toch een sterk vertrouwen dat een dergelijk systeem in real-time zal werken op kleine, goedkope computersystemen. Ter vergelijking, in (Luo, 2018) worden verschillende *motion-prediction* systemen vergeleken die een snelheden halen van 9ms tot 65ms op enorm krachtige en prijzige (€5200) opstellingen.

7. Discussie

Zoals eerder vermeld, is een groot deel van het verlies in accuraatheid van de netwerken te wijten aan de sterke grenzen die opgesteld zijn bij het definiëren van gevaar. Daarom zou het in toekomstige onderzoeken een goed idee zijn om in de plaats van een binaire classificatie van "gevaarlijk" of "veilig" een eerder regressieve classificatie te gebruiken. Dit houdt in dat het neurale netwerk kan voorspellen niet enkel voorspelt of een situatie al dan niet gevaarlijk zal zijn, maar ook hoe gevaarlijk deze zal zijn. De eerste stap om dit soort systeem te ontwikkelen, is het bedenken van een schaal van gevaar. Dit kan bijvoorbeeld de tijd tot impact, het risico op impact, de snelheid van het voertuig, ... in beschouwing nemen.

Verdere uitbreidingen zijn het uittesten van deze systemen op zeer lichte computersystemen zoals een Raspberry Pi. Ook zou het interessant zijn om te werken met goedkopere LiDAR alternatieven. Wegens de populariteit van LiDAR systemen, daalt de gemiddelde prijs van de sensoren drastisch jaar na jaar (Neal, 2018). Toch zou het een interessante uitbreiding zijn om dit soort systeem uit te proberen met goedkopere *single-beam* LiDAR sensoren die tegenwoordig al aan een zeer scherpe prijs te vinden zijn. Dit werd al gesimuleerd door de NuScenes LiDAR scans om te vormen naar tweedimensionale polaire roosters met een veel lagere resolutie.

8. Conclusie

Er werd in deze paper aangetoond dat het mogelijk is om een zeer licht systeem te ontwikkelen dat kan voorspellen of er al dan niet gevaar op een botsing zal optreden in de nabije toekomst. Deze voorspellingen konden meer dan de helft van de gevaarlijke situaties correct voorspellen zonder een grote hoeveelheid valse alarmen. Een dergelijk concept zou veel haalbaarder zijn voor fiets-gemonteerde systemen dat de huidige veiligheidssystemen ontworpen voor voertuigen.

References

- Bianchi, F. M. (2017). *Recurrent Neural Networks for Short-Term Load Forecasting*. Springer International Publishing.
- Buchholtz, K. &. (2020). An evaluation of bicycle-specific agility and reaction times in mountain bikers and road cyclists. *South African Journal of Sports Medicine*, 1-5.
- Caesar, H. &. (2019). nusenes: A multimodal dataset for autonomous driving. arXiv.
- Djuric, N. &. -C.-H. (2020). Uncertainty-aware Short-term Motion Prediction of Traffic Actors for Autonomous Driving. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, (pp. 2084-2093).
- Hahner, M. &. (2020). Quantifying Data Augmentation for LiDAR based 3D Object Detection. <https://github.com/nutonomy/second.pytorch>.
- Kim, J.-K. &. (2007). Bicyclist injury severities in bicycle-motor vehicle accidents. Accident; analysis and prevention. *Accident; analysis and prevention, vol 39*, 238-51.
- Luo, W. &. (2018). Fast and Furious: Real Time End-to-End 3D Detection, Tracking and Motion Forecasting with a Single Convolutional Net. *IEEE/CVF Conference on Computer Vision and Pattern Recognition,,* (pp. 3569-3577).
- Neal, A. (2018, April 24). LiDAR vs. RADAR. Fierce Electronics.
- Schreiber, M. &. (2020). Dynamic occupancy grid mapping with recurrent neural networks. *CoRR*.
- Wang, P. &. -Y. (2017). Vehicle collision prediction at intersections based on comparison of minimal distance between vehicles and dynamic thresholds. *IET Intelligent Transport Systems, vol. 11, issue 10*, 676 -684.