2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

A Safety-Adapted Loss for Pedestrian Detection in Autonomous Driving

Maria Lyssenko*‡, Piyush Pimplikar*, Maarten Bieshaar*, Farzad Nozarian**, Rudolph Triebel†§
* Robert Bosch GmbH, Corporate Research, Germany, firstname.lastname@de.bosch.com
‡ Technical University of Munich, Germany, firstname.lastname@tum.de
** German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany, farzad.nozarian@dfki.de
† German Aerospace Center (DLR), Wessling, Germany, rudolph.triebel@dlr.de
§ Karlsruhe Institute of Technology, Germany, rudolph.triebel@kit.edu

***Abstract*— In safety-critical domains like autonomous driving (AD), errors by the object detector may endanger pedestrians and other vulnerable road users (VRU). As raw evaluation metrics are not an adequate safety indicator, recent works leverage domain knowledge to identify safety-relevant VRU, and to back-annotate the criticality of the interaction to the object detector. However, those approaches do not consider the safety factor in the deep neural network (DNN) training process. Thus, state-of-the-art DNN penalize all misdetections equally irrespective of their importance for the safe driving task. Hence, to mitigate the occurrence of safety-critical failure cases like false negatives, a safety-aware training strategy is needed to enhance the detection performance for critical pedestrians. In this paper, we propose a novel, safety-adapted loss variation that leverages the estimated per-pedestrian criticality during training. Therefore, we exploit the reachable set-based time-to-collision (TTC_{RSB}) metric from the motion domain along with distance information to account for the worst-case threat. Our evaluation results using RetinaNet and FCOS on the nuScenes dataset demonstrate that training the models with our safety-adapted loss function mitigates the misdetection of safety-critical pedestrians with robust performance for the general case, *i.e.*, safety-irrelevant pedestrians.**

I. Introduction

Whenever autonomous mobile robots or autonomous vehicles (AV) operate in dynamic and highly complex environments ensuring correct and reliable detection of vulnerable road users (VRU) becomes vital. In respect thereof, current training and evaluation approaches of state-of-the-art object detectors have been extensively studied as an enabling technology, attributed to the remarkable success in camera-based perception [1].

In contrast to non-safety critical computer vision applications like, *e.g.*, parking occupancy monitoring in car parks or tallying people for waiting time analytics [2], cases of failed detections (so-called *false negatives*) in autonomous driving (AD) may entail a safety risk [3], [4]. Hence, to prevent collisions, it is of utmost importance to consider the safety factor in the training process to assure accurate perception capabilities for relevant VRU.

Let us consider a pedestrian detector applied to a crowded urban scene. Here, perception failures in vicinity of the AV are *safety-critical* as misdetected pedestrians may pose an imminent collision risk (in, *e.g.*, street crossing scenarios from Fig. 1 highlighted by the red bounding box), whereas distant errors do not directly affect the safe driving task (*e.g.*, orange bounding boxes). Subsequently, the promotion of safe driving behavior necessitates *(i)* a comprehensive identification of all safety-critical pedestrians in the urban scene [5], [6] and

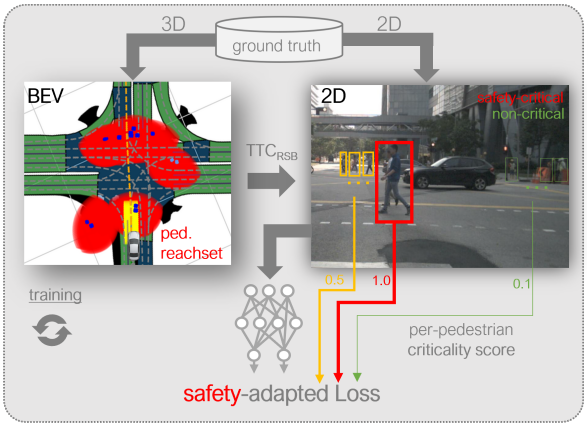

Fig. 1: We propose the safety-adapted focal loss to enhance the detection performance for safety-critical pedestrians. The loss diligently exploits distance information and a threat metric from the motion domain (BEV) to consider the criticality of individual pedestrians during the training of an one-stage 2D object detector (vision domain). Concretely, we leverage reachable set-based time-to-collision (TTC_{RSB}), defined as the intersection between the AV's (yellow) and pedestrians' (red) reachable sets, to derive the *per-pedestrian criticality* for each pedestrian in the scene. We link the *criticality scores* of the detections to the *safety-adapted loss* function to dynamically adjust the loss contribution of safety-critical pedestrians.

(ii) the guarantee of impeccable detection results specifically for these pedestrians at risk [7], [8]. Therefore, to tackle the underspecification of aggregated, purely vision-based metrics [9], recent works by Wolf *et al.* [10], Bansal *et al.* [11], and Lyssenko *et al.* [12] encompass a notion of criticality into the evaluation of the utilized object detector. As an example, in our investigated pedestrian detection use case in [12], we employ the reachable set-based time-to-collision (TTC_{RSB}) determining the earliest point in time when a collision with the AV may occur. Here, our evaluation has encountered several safety-critical misdetections within a significant number of pedestrian sequences. Consequently, the question arises:

How can we remedy potentially safety-critical pedestrian misdetections?

As our main contribution, we propose a novel, safety-adapted loss function that effectively exploits the criticality of individual pedestrians during training as illustrated in Fig. 1. Therefore, we *(i)* derive our per-pedestrian criticality score from the motion

domain and *(ii)*, we incorporate the criticality into the focal loss to dynamically adjust the loss contribution. Intuitively, by factoring in the criticality in the safety-adapted loss, we amplify the loss contribution to focus the DNN on safety-critical pedestrians. We provide an experimental evaluation of our safety-adapted loss utilizing the nuScenes dataset [13] and two state-of-the-art object detectors (RetinaNet [14] and FCOS [1]). Our results demonstrate, that we successfully decrease the misdetection of safety-critical pedestrians with robust overall detection performance.

The remainder of the paper is structured as follows. First, we review the related work in Sec. II, before introducing our methodology for constructing the safety-adapted loss function in Sec. III. Thereafter, we provide the experiment setup in Sec. IV, followed by the experimental results in Sec. V.

II. Related Work

A. Safety-Awareness in Automated Driving

Commonly used evaluation metrics like intersection-over-union (IoU), mean average precision (mAP), or recall are widespread because they are not specific to a particular task and allow for meaningful comparisons across different benchmarks [15]. However, due to the safety-agnostic character of those metrics there is no adequate assessment whether the perception function guarantees sufficient detection performance when deployed in the safety-critical AD domain [8].

The approaches by Wolf *et al*. [10], Bansal *et al*. [11], and Ceccarelli *et al*. [16] extend pure distance-based domain knowledge in metrics design [17], [18]. They argue that a distance-based, *potential* collision risk does not consider the dynamics and the criticality of the interaction. Therefore, the authors propose a definition of the *imminent* collision risk implementing threat metrics from the AD domain such as TTC to account for a safety-indicator in aggregated evaluation measures. However, the authors rely on simplified motion models assuming constant velocity vectors and heading charted over a time horizon.

Considering all possible, worst-case states an agent can reach within a time interval, recent works by Topan *et al*. [19], [20] and Lyssenko *et al*. [12] implement the identification of safety-relevant agents using dynamic-aware perception zones derived from reachability analyses. The work in [19] utilizes a Hamilton-Jacobi (HJ) reachability to construct a sound safety zone around the AV that shall encompass all safety-relevant agents. In contrast, Lyssenko *et al*. [12] focus on functional scenarios, *i.e*., individual interaction with the AV. The work assumes that the AV is lane bound and implement a map-constrained calculation of the AV's reachable set leveraging motion models based on differential inclusions. To account for uncertainties of future motion and the resulting worst-case criticality assessment, we utilize the framework proposed by Lyssenko *et al*.[12] to derive the TTC_{RSB} for potentially dangerous interactions between pedestrians and AV.

B. Significance of Loss Functions

One of the essences of the object detection task revolves around the significance of loss functions [21]. Therefore, recent progress has shifted from generic loss functions such as binary-cross-entropy (BCE) towards novel alternatives like the focal loss [14], [1], to reduce the importance of well-classified samples.

Further work by Li *et al*.[22] extend the focal loss using a category-relevant, dynamic modulating factor to increase the impact of rare categories. This also motivates our work to include a criticality component in the focal loss to magnify the loss contribution for individual pedestrians under risk.

Besides a conceptual proposal that emphasizes mitigation strategies in the network construction process [7], to the best of our knowledge, the first implementation of a safety-adapted regression loss is presented by Liao *et al*. [23] for 3D object detection. The authors additively combine the Smooth-L_1 with a safety loss component to minimize the discrepancy between prediction and ground truth (GT) volume ratios. Although the work proposes a distance ratio to additionally quantify the misalignment between GT and prediction *w.r.t*. the criticality, spatial properties are not further explored in the loss construction itself.

As motivated by Abrecht *et al*. [24], we want to tackle safety-agnosticism in the loss construction process to mitigate the occurrence of safety-critical false negatives. Therefore, we leverage knowledge from the application domain to propose a variant of the safety-adapted focal loss.

III. Methodology

In this section, we formulate our novel safety-adapted focal loss. To mitigate the occurrence of critical misdetections, we construct our loss based on the dynamic-aware per-pedestrian criticality that encompasses the worst-case collision risk. In Sec. III-A, we introduce the collision risk on the basis of TTC_{RSB} from reachability analyses and present the combined per-pedestrian criticality in Sec. III-B. We motivate the design of our safety-adapted loss in Sec. III-C.

A. Collision Risk from Reachable Sets

The adequate assessment of an interaction's criticality between a pedestrian and the AV like in Fig. 2, requires a prediction on how the situation may evolve in the future, given the current state and the underlying motion models for the AV and the pedestrians, respectively.

In the following, we employ the reachability framework from previous work [12], [25] to estimate the per-pedestrian criticality from the AV's perspective. Therefore, we exploit differential inclusion-based motion models to provide a safe overapproximation of the possible future states of the pedestrian and the AV, respectively [26].

We leverage those models to acquire a clear specification of the objects' expected movement over a certain amount of time to calculate the so-called reachable set, *i.e*., a set of all possible future states the object could reach irrespective of the probability. Thereby, we calculate the reachable sets from Fig. 2 for each pedestrian $R_{ped,i}$ (red) and our automated vehicle R_{AV} (yellow). For the motion model definition, we utilize a constant acceleration model for the pedestrians and a constant velocity model from [12] for the AV. In mathematical terms: For each

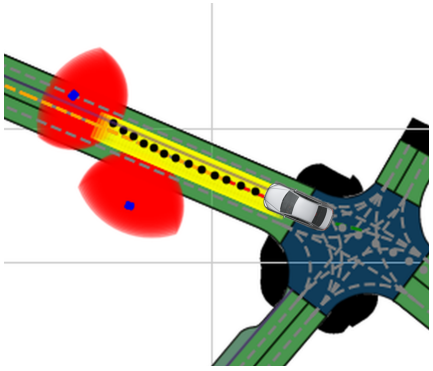

Fig. 2: Exemplary illustration of a safety-critical interaction at time τ between an AV and pedestrians at risk (blue markers) and the respective reachable sets $R_{AV}(\tau)$ and $R_{ped,i}(\tau)$ for a given AV trajectory (black dots). The intersection of the reachable sets emphasizes the collision risk with $TTC_{RSB} < \infty$, *i.e.*, $R_{AV}(\tau) \cap R_{ped}(\tau) \neq \emptyset$. To calculate $R_{AV}(\tau)$ *w.r.t.* the planned driving corridor [25], we require a sequence of the current (red) and successive centerlines (orange) as the input to the reachability framework from [12] to define our road.

pedestrian $i \in \{1,\dots,N\}$ in a scene with N pedestrians, a given current system state $\boldsymbol{x}_i(\tau)$ at time $\tau \in \mathbb{R}^+$, its initialization for each $\boldsymbol{x}_i(0)$ (denoted by the initial position, velocity, and acceleration), and the constant acceleration motion model F_a in terms of differential inclusions as $\dot{\boldsymbol{x}}_i(\tau) \in F_a(\boldsymbol{x}_i(\tau))$, we can calculate the resulting reachable set

$$R_{ped,i}(\tau) = \left\{ \boldsymbol{x}_{u,i}(\tau) | \boldsymbol{x}_{u,i}(\tau) = \boldsymbol{x}_i(0) + \int_0^\tau u_i(t)dt \right\}, \quad (1)$$

$$\text{where } \forall t_1 \leq \tau : u_i(t_1) \in F_a(\boldsymbol{x}_i(t_1)). \quad (2)$$

Eq. 2 means that $R_{ped,i}(\tau)$ contains all states $\boldsymbol{x}_{u,i}(\tau)$ that start at $\boldsymbol{x}_i(0)$ and can be reached by the trajectories $u_i(t)$ that are contained by $F_a(\boldsymbol{x}_i(t))$ with $t \in [0,\tau]$. Please note, we perform the $R_{AV}(\tau)$ calculation in accordance to Eq. 2 employing the constant velocity model $F_v(\boldsymbol{x}_{AV}(t))$.

To estimate the criticality between each pedestrian i and the AV, we use time-to-collision (TTC) as our threat metric, which quantifies the earliest point in time

$$TTC_{RSB,i} = \min\{\tau \mid R_{AV}(\tau) \cap R_{ped,i}(\tau) \neq \emptyset\} \quad (3)$$

when two reachable sets intersect, *i.e.*, the first point in time where a collision *may* happen. Following Schneider *et al.* [25], we exploit a TTC formulation based on reachable sets to extend the current state-of-the-art TTC formulation that utilizes point estimates without uncertainties [27]. Please note, there are current deep learning approaches to predict the TTC from mono-camera input utilizing, *e.g.*, optical flow [28]. However, we opt to use precise ground truth information to facilitate white-box methods to argue safety [8].

B. Per-Pedestrian Criticality

The initial work on $TTC_{RSB,i}$ in [12] exploits reachable sets to identify pedestrians at risk of an imminent collision below a TTC threshold, *i.e.*, $TTC_{RSB,i} < TTC_{crit}$. However, the study revealed that the employment of the constant velocity model for R_{AV} calculation may produce a *blind spot*, *i.e.*, an insensitivity for $TTC_{RSB,i} > TTC_{crit}$.

Concretely, given a low AV velocity, the reachability analyses produces a R_{AV} of a small spatial extent and thus, a non-critical $TTC_{RSB,i}$ although the pedestrian is in the direct vicinity of the AV. Therefore, we additionally inject distance information to account for the potential collision risk for non-critical pedestrians *w.r.t.* $TTC_{RSB,i}$.

Based on this underlying idea, we compose our criticality weighting for individual pedestrians κ_i by means of *(i)* $TTC_{RSB,i}$ that accounts for the collision criticality ($\kappa_{c,i}$) considering the uncertainty-aware dynamics of the interaction, and *(ii)* the distance between the pedestrian and AV to reflect the distance criticality ($\kappa_{d,i}$) irrespective of the motion model.

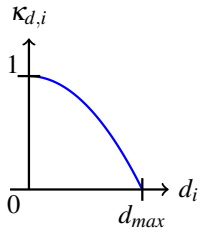

Fig. 3: Downward parabola to model the distance criticality $\kappa_{d,i}$ over the distance d_i to a pedestrian i.

Let us now focus on the implementation of $\kappa_i(\kappa_{c,i}, \kappa_{d,i})$, where we want to design $\kappa_i \in [0,1]$, *i.e.*, $\kappa_i = 1$ represents a pedestrian of highest relevance for the driving task. Inspired by the work of Ceccarelli *et al.* [16], we utilize the downward parabola from Fig. 3 that passes through the points $(0,1)$ and $(d_{max},0)$, to describe $\kappa_{d,i}$ over distance d_i. Please note, d_{max} describes the distance up to which we consider a pedestrian as safety-relevant for the driving task. Further, we leverage the non-linear decrease of $\kappa_{d,i}$

$$\kappa_{d,i}(d_i) = -\frac{1}{d_{max}^2} d_i^2 + 1, \text{ where } d_i \in [0, d_{max}] \wedge \kappa_{d,i} \in [0,1] \quad (4)$$

to achieve a slow decrease in the distance criticality for $d_i \to 0$, *i.e.*, for pedestrians close to the AV. Thus, for distant pedestrians with $d_i \to d_{max}$, we estimate $\kappa_{d,i} \to 0$. To this end, we apply Eq. 4 for $\kappa_{c,i} \in [0,1]$, respectively, to estimate the collision criticality according to the non-linear decrease *w.r.t.* the time t. Here, we use TTC_{max} as the time threshold leading to $TTC_{RSB,i} \to TTC_{max}$ and $\kappa_{c,i} \to 0$ for distant pedestrians.

For the composed per-pedestrian criticality κ_i

$$\kappa_i = \frac{1}{3}(2\kappa_{c,i} + \kappa_{d,i}) \quad (5)$$

$$\text{where } \kappa_{c,i}, \kappa_{d_i}, \kappa_i \in [0,1], \quad (6)$$

we implement a double weighting of $\kappa_{c,i}$ in the formula [10] as it includes a dynamics-aware criticality estimate and thus, a superior measure of the collision risk.

C. Safety-adapted Focal Loss

The focal loss (*cf.* Eq. 7) is commonly employed in object detectors to mitigate the foreground-background imbalance [14]. Thereby, the key idea of the loss is to re-balance the loss contribution of easy samples, *i.e.*, decrease their importance in the training process. As described in [14], the focal loss

$$FL(p_i) = -\alpha(1-p_i)^\gamma \log(p_i) \quad (7)$$

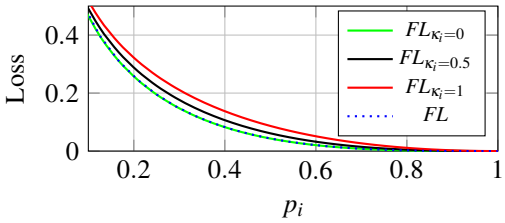

Fig. 4: Illustration of the focal loss FL (with $\gamma = 2, \alpha = 0.25$) and the safety-adapted focal loss $FL_{\kappa,i}$ *w.r.t.* the pedestrians' criticality κ_i. For $\kappa_i = 0$ (non-critical pedestrians), we preserve the properties of the FL, *i.e.*, $FL_{\kappa,i} \rightarrow FL$. For critical pedestrians with $\kappa \rightarrow 1$, we naturally magnify the loss contribution, where $FL_{\kappa \rightarrow 1} \geq FL$.

implements an extension of the generic BCE loss by utilizing a weighting factor α and a modulating factor $(1-p_i)^{\gamma}$. Given the definition, the focusing parameter γ determines the properties of the FL and down-weights easy samples *w.r.t.* to the predicted object's class probability $p_i \in [0,1]$. Concretely, a higher γ extends the probability range when a sample is considered easy and thus, lowers the loss contribution for those well-classified samples accordingly. However, the current loss determination is solely dependent on p_i and the hyperparamers (α, γ) irrespective of the objects' criticality. To tackle the importance of objects, Li *et al*. [22] emphasize a category-dependent focusing factor for a long-tail distribution, *i.e.*, in our case the imbalance between critical and non-critical pedestrians (see Fig. 6).

Inspired by Li *et al*. [22], we want to inject the criticality into the loss but on an instance level, *i.e.*, for individual pedestrians, as for safety considerations there should be a discrimination between safety-relevant and safety-irrelevant objects within a category. In our work, we leverage the criticality κ_i from Sec. III-B to amplify the loss contribution for critical pedestrians with $\kappa \rightarrow 1$ in our safety-adapted FL, *i.e.*,

$$FL_{\kappa_i}(p_i, \kappa_i) = -\alpha(1-p_i)^{(\gamma-\kappa_i)} log(p_i). \tag{8}$$

Given the properties of the FL, a larger $\gamma > 2$ is used for a severe positive-negative imbalance, which will result in a sacrifice *w.r.t.* to the samples' loss contribution in the training process. This limits the performance for rare samples, *i.e.*, safety-critical pedestrians. With our instance-based criticality weighting κ_i, we want to counteract the diminishing loss contribution for critical pedestrians and propose the adaptation of the focusing parameter to $(\gamma\text{-}\kappa_i)$ with $\gamma = 2$ [14] as illustrated in Fig. 4.

Thereby, we *(i)* dynamically change the loss contribution for $\kappa_i > 0$ *w.r.t.* a pedestrian's criticality in the training process, and *(ii)* we maintain the properties of the FL for non-critical pedestrians, *i.e.*, with $\kappa_i \rightarrow 0$ we obtain $FL_{\kappa_i} \rightarrow FL$. Please note, for our weighting we derive κ_i directly from the motion domain, *i.e.*, no extensive hyperparameter search is required.

IV. Experiment Setup

In the following, we describe the setup to evaluate our novel safety-adapted loss. In Sec. IV-A and Sec. IV-B, we introduce the utilized datasets, and the data cleaning procedure, respectively. In Sec. IV-C, we elaborate on the training protocols for the employed pedestrian detectors.

A. Datasets from nuTonomy

For our experiments, we employ two datasets from nuTonomy: nuImages and nuScenes [13]. We use the 2D nuImages dataset with its precise bounding box annotations for the pedestrian object detector's initial pre-training.

As the criticality estimation necessitates domain information, we require a 3D dataset with position and velocity information of the pedestrians along with map information for the AV to calculate the $TTC_{RSB,i}$ and distance information, respectively. Therefore, we utilize nuScenes as it provides data from the entire sensor suite of an AV for 1000 scenes. We perform the training and evaluation on the splits as defined in the *nuScenes-devkit* [29].

Please note, for our experiments, we utilize the images from the front camera only, and the corresponding lidar point clouds of the scene that are matched over a scene `token`. However, despite the rich annotations, the nuScenes dataset contains only 3D bounding boxes. Therefore, we project the cuboids' coordinates onto the camera pixel grid to retrieve 2D pedestrian annotations using the helper function `get_2D_boxes()`.

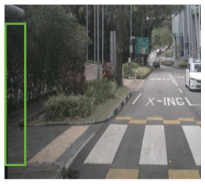

Fig. 5: Invalid 2D bounding box due to nuScenes' annotation policy.

B. Data Curation on nuScenes

The nuScenes labelling policy discards object boxes without any lidar and radar points to filter out temporarily fully occluded objects [30]. However, there are still occurrences of false positive annotations as illustrated in Fig. 5. As the `get_2D_boxes()` function projects the cuboids into the frames of all cameras, bounding box projections from the left and right cameras might appear in the relevant front camera frame. To mitigate such artifacts, we utilize the pedestrians' position information from the motion domain. Thereby, for each projected box, we determine whether the center of its cuboid falls within the AV's physical field-of-view of the front camera, which is 70°. In the case of a missing correspondence, we associate the cuboid with one of the side cameras and discard the respective 2D annotation from the front camera in the training and evaluation phase.

C. Training Protocol: Pedestrian Detectors

For our pedestrian detectors, we implement the RetinaNet [14] and the FCOS [1] using PyTorch [31] and employ the following pre-training protocols on nuImages.

RetinaNet: In the implementation from [32], we utilize the ResNet-50 backbone, the Adam optimizer with a learning rate of $1e^{-5}$, the reduce-on-loss-plateau scheduler (patience=3), and we train our model for 200 epochs using a batch size of 16. We obtain an AP50 of 0.31 for the pedestrian class on the nuImages validation split.

FCOS: We follow the original paper implementation from [33] with the ResNet-50 backbone that is trained for 42 epochs using the batch size of 16. During training, we employ the stochastic gradient descent optimizer with an initial learning rate of $1e^{-3}$. Furthermore, we apply a multi-step learning rate decay with a linear warm-up. Here, we obtain an AP50 of 0.48 for the pedestrian class.

Safety-adapted training: For the implementation of the safety-adapted loss from Sec. III-C for both pedestrian detectors, we use the respective models pre-trained on nuImages as they indicated a reasonable performance for the pedestrian class. With the safety-adapted loss, we train the models on the nuScenes' training split until the loss on the validation set converges (≈ 4 epochs). More specifically, we leverage the estimated criticality from Sec. III-B to dynamically adapt the modulating factor for the pedestrian class. For other categories like cars and the background class, we set $\kappa = 0$ to maintain the the properties of the focal loss.

V. Experimental Results

In Sec. V-A, we present the evaluation of our novel safety-adapted focal loss for RetinaNet and FCOS *w.r.t.* the focal loss baseline, and investigate the impact on detection capabilities for pedestrians of a different criticality. Further, in Sec. V-B we relate the safety-adapted loss to the pedestrians' detection easiness, and in Sec. V-C, we analyze how the design of the per-pedestrian criticality affects the safety-critical performance.

A. Safety-Adapted Loss Evaluation

We start our evaluation by defining three zones (with the corresponding pedestrian count) that encompass the critical (159), potentially critical (1126) and non-critical pedestrians (3371) as illustrated in the heatmap in Fig. 6. For a distance $>$ $40\,m$ we have additional 3025 non-critical pedestrian instances that are not visualized in Fig. 6.

Depending on the TTC_{RSB} and the distance for individual pedestrians, each cell summarizes the respective count in the nuScenes validation set. Please note, that the lower right part of the heatmap does not contain any samples as the AV's velocity is thresholded by the urban speed limit of up to 30 *mph* ($\approx 13,3 ms^{-1}$). Hence, the speed limit lower-bounds the TTC_{RSB} that is feasible for a given distance.

Consequently, for the given speedlimit a braking time of $1.7s$ might be required to avoid a collision [12]. Therefore, we define the safety-critical zone $\boldsymbol{C}$ by *(i)* $TTC_{crit} = 1.7\,s$, and *(ii)* a critical distance of $\text{dist}_{crit} = 20\,m$ that contains roughly 2.1% of all pedestrians. The potentially safety-critical zone (PC) is lower-bounded by TTC_{crit} with a distance up to d_{crit} with 14.8% of all pedestrians, and the non-critical zone (NC) contains the remaining pedestrians of the validation set. Given the discussed blind spot for a low AV's velocity in Sec. III-B (high TTC_{RSB} below d_{crit}), we explicitly consider the potentially critical zone *PC* in our evaluation as the AV still may accelerate to the speed limit and thus, shift the pedestrians into the safety-critical zone $\boldsymbol{C}$.

In Tab. I, we outline the effectiveness of our approach for the safety-critical zone $\boldsymbol{C}$. Therefore, we compare the safety-adapted loss FL_κ to the baseline focal losses $FL_{\gamma=1}$ and $FL_{\gamma=2}$.

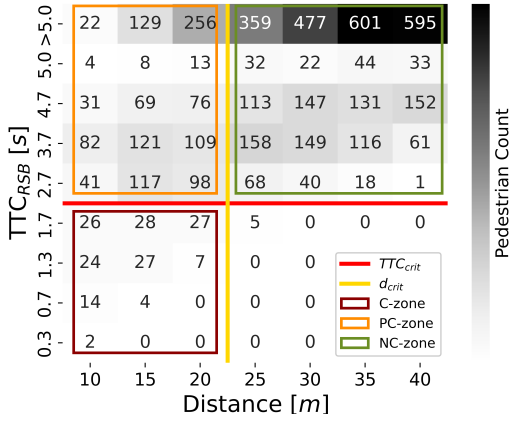

Fig. 6: The heatmap illustrates the pedestrian count *w.r.t.* the corresponding TTC_{RSB} and distance values for the safety-critical $\boldsymbol{C}$ (159, *dark red*), potentially critical *PC* (1126, *orange*) and non-critical (3317 for $d > 20\,m$, *green*) zone. The safety-critical zone is defined by TTC_{crit} and dist_{crit}.

TABLE I: Performance of RetinaNet and FCOS on nuScenes utilizing the baseline focal loss $FL_{\gamma=2}$ and the safety-adapted focal loss FL_κ. AP^S, AP^M, and AP^L show AP^{50} scores for *all* pedestrians conditioned on the bounding box diagonals, *i.e.*, $< 150\,px$ (small), 151-$350\,px$ (mid), and $> 350\,px$ (large). $\text{Recall}^{\boldsymbol{C}}$ and Recall^{NC} define the recall for critical and non-critical pedestrians.

Model	Method	AP^{50}	AP^S	AP^M	AP^L	$\text{Recall}^{\boldsymbol{C}}$	Recall^{PC}	Recall^{NC}	Precision
RetinaNet−50	$FL_{\gamma=2}$	0.441	0.366	0.667	0.656	0.881	0.790	0.441	0.782
	$FL_{\gamma=1}$	0.429	0.347	0.691	0.746	0.899	0.775	0.411	0.830
	$\boldsymbol{FL_\kappa}$	0.440	0.356	**0.729**	0.656	**0.906**	**0.803**	0.433	0.778
FCOS−50	$FL_{\gamma=2}$	0.476	0.424	0.616	0.615	0.918	0.879	0.649	0.316
	$FL_{\gamma=1}$	0.457	0.412	0.667	0.521	0.830	0.796	0.535	0.320
	$\boldsymbol{FL_\kappa}$	0.474	0.422	0.663	0.594	**0.950**	0.878	0.644	**0.322**

The table demonstrates that in comparison to FL_κ, the "vanilla" reduction of the focusing parameter to $\gamma = 1$ decreases the recall scores ($\text{Recall}^{\boldsymbol{C}}$) for RetinaNet and FCOS by 0.7% and 12%, respectively. Similarly for *PC*, we denote a performance decline for both models by 0.8% and 8.2% when evaluated with $FL_{\gamma=1}$. Thereby, from the results we can conclude that a naive decrease in γ for all samples (and consequently a higher loss contribution for all samples irrespective the criticality) leads to a diminished recall in $\boldsymbol{C}$ and *PC*, *i.e.*, a higher number of (potentially) critical misdetections.

Thus, the results reflect that the dynamic adjustment of the focusing parameter $(\gamma - \kappa)$ achieves superior sensitivity in $\boldsymbol{C}$ in comparison to $FL_{\gamma=1}$ and $FL_{\gamma=2}$. Particularly for RetinaNet and FCOS in $\boldsymbol{C}$, our approach improves the $FL_{\gamma=2}$ recall baseline by 2.5% and 3.2%, respectively. For *PC*, our evaluation demonstrates a recall incline by 1.3% for the RetinaNet and stable performance for FCOS.

Conclusively, the results from Tab. I emphasize that FL_κ can handle the severe imbalance between critical, potentially critical, and non-critical pedestrians with a negligibly small performance degradation of $< 1.0\%$ in the overall performance (*cf.* AP^{50}, AP^S, AP^M, AP^L, and Precision in Tab. I). Please note, that for our zone-based evaluation we primarily employ

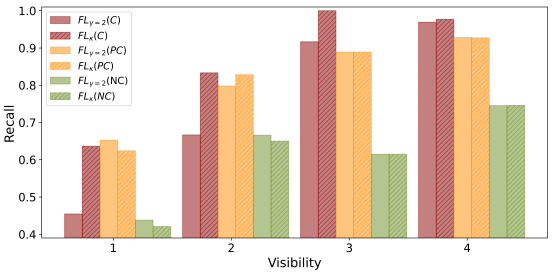

Fig. 7: The bar plot depicts the recall performance of FCOS *w.r.t.* the visibility from 1 (hardest) to 4 (easiest) for critical, potentially critical, and non-critical pedestrians evaluated with FL_{κ} and $FL_{\gamma=2}$.

recall scores as we require physical properties to identify pedestrians as critical, potentially critical, or non-critical. As false positive detections lack associated 3D ground truth, we can not calculate the precision for the three zones trivially. We leave this for future work. However, in the overall precision and AP^{50} scores for RetinaNet, we observe only a minor decrease of up to 0.4% across all criticality zones. Specifically for FCOS, there is even a further improvement in precision by 0.6%.

B. Ablation Study: Criticality and Learning Difficulty

In this ablation study, we investigate to which extent the safety-adaptation of the loss relates to the detection's easiness or hardness. In other words, we want to ensure that our model trained with FL_{κ} does not only learn to detect "easy" but critical pedestrians, *e.g.*, unoccluded pedestrians in the direct vicinity those that are easily visible. Among numerous definitions [34] to define detection and sample difficulty, we use the annotated visibility from nuScenes dataset as a simple proxy.

We conduct the detailed analyses on the FCOS as it outperformed RetinaNet by 3.4% and 4.4% *w.r.t.* AP^{50} and Recall$^{\boldsymbol{C}}$. The bar plot in Fig. 7 depicts the recall values for the pedestrians within the zones of different criticality from Fig. 6, *i.e.*, for the zones $\boldsymbol{C}$, *PC* and *NC*, evaluated with FL_{κ} and the $FL_{\gamma=2}$, respectively, for different partitions of pedestrian visibilities (4 bins representing a decreasing difficulty) [13] up to $40m$. As expected, for all categories the distribution shows an increasing recall trend with a higher visibility (easier samples). The bar plot also illustrates approximately equal recall scores for potentially critical and non-critical pedestrians from *PC* and *NC* for higher visibilities. Given our definition of FL_{κ}, we would expect that behaviour as we design $FL_{\kappa} \rightarrow FL_{\gamma=2}$ for pedestrians of lower criticality with $\kappa \rightarrow 0$. For lower visibilities (bin 1 and 2), the distribution shows that except for *PC* our safety-adapted focal loss F_{κ} outperforms the baseline $FL_{\gamma=2}$. Particularly, for the critical zone, we denote a recall increase for all visibilities, which indicates that we are able to mitigate false negatives within partitions of different difficulty.

C. Ablation Study: TTC vs. Distance

In our second ablation study, we evaluate the impact of individual components of the composed per-pedestrian criticality on the performance. Therefore, in Tbl. II, we decouple the distance criticality κ_d and the collision criticality κ_c from Sec. III-B into individual losses (FL_{κ_d} and FL_{κ_c}) and compare the results to the baseline FL_{κ}. As in Sec. V-A, we perform the recall evaluation for our three zones of different criticality.

TABLE II: Extension of Tbl. I to illustrate the ablation study on the composed criticality κ *w.r.t.* collision criticality κ_c and distance criticality κ_d.

Model	Method	AP^{50}	AP^{S}	AP^{M}	AP^{L}	Recall$^{\boldsymbol{C}}$	Recall^{PC}	Recall^{NC}	Precision
RetinaNet−50	FL_{κ}	0.440	0.356	0.729	0.689	0.906	0.803	0.492	**0.778**
	FL_{κ_c}	**0.445**	**0.358**	**0.741**	0.689	**0.918**	**0.810**	**0.505**	0.741
	FL_{κ_d}	0.441	0.353	0.724	**0.796**	0.906	0.800	0.503	0.721
FCOS−50	FL_{κ}	**0.474**	**0.422**	0.663	0.594	**0.950**	0.879	0.669	**0.322**
	FL_{κ_c}	0.467	0.418	**0.690**	0.617	0.925	0.881	**0.670**	0.315
	FL_{κ_d}	0.463	0.416	0.681	**0.637**	**0.950**	**0.888**	0.669	0.291

The results show that there is a different trend for our evaluated losses *w.r.t.* a particular criticality. As we can see, FL_{κ_c} outperforms the baseline FL_{κ} among almost all categories for the RetinaNet. It shows also superior results to the pure distance criticality FL_{κ_d} except for AP^{L} with a negligibly small decrease of only 0.7%. However, for FCOS the performance scores do not reflect a sound indication for a particular criticality choice. From the recall evaluation of the three zones, we can derive that for zone $\boldsymbol{C}$ pure distance information was *(i)* sufficient to reach the performance from the baseline, and *(ii)* able to achieve a higher recall for *PC* by almost 1%, while maintaining stable results within *NC*.

In conclusion, although our TTC_{RSB}-based criticality from Sec. III-A promotes the identification of safety-critical pedestrian *w.r.t.* dynamic properties, our ablation study has shown that depending on the model, a simple criticality measure like distance may be employed as a reasonable proxy for relevance during training.

VI. Conclusion and Future Work

Safety-agnostic training is a great safety concern for perception in AD. Therefore, this work presents a novel, safety-adapted focal loss that leverages domain knowledge (*i.e.*, per-pedestrian criticality) during training to mitigate the occurrence of critical misdetections (false negatives). We evaluate the novel loss on the safety-critical zone defined by a $\text{TTC}_{crit} < 1.7s$ and $\text{d}_{crit} < 20m$ and show that for RetinaNet-50 and FCOS-50 we achieve a recall increase of 2.5% and 3.2%, respectively. Supplementary, we demonstrate that the novel loss maintains stable overall performance for pedestrians outside the safety-critical zone. This, in particular, enables the employment of the safety-adapted focal loss for AD applications as the initial concept provides valid and promising results.

Up to now, we have only considered the mitigation of false negatives, but from a safety perspective, false positives are also of great concern. An approach to determine the criticality of falsely detected pedestrians should be included in future work. Further, the extension of the method to other classes and loss functions shall be considered. For instance, we plan to extend the safety-adapted loss to the regression task to improve the detection quality of critical pedestrians.

References

[1] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," 2019.

[2] "viso.ai," 2023, accessed on September 14, 2023. [Online]. Available: https://viso.ai/applications/computer-vision-applications/

[3] T. W. Post, "How a robotaxi crash got cruise's self-driving cars pulled from californian roads," https://www.washingtonpost.com/technology/2023/10/28/robotaxi-cruise-crash-driverless-car-san-francisco/, 23.10.2023, accessed: 17.11.20232.

[4] N. T. S. Board, "Collision between vehicle controlled by developmental automated driving system and pedestrian," https://www.ntsb.gov/investigations/Pages/HWY18MH010.aspx, 18.03.2018, accessed: 17.11.2023.

[5] C. Bürkle, F. Oboril, J. Jarquin, F. Pasch, and K.-U. Scholl, "Safe perception: On relevance of objects for vehicle safety," *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 3957–3964, 2021.

[6] K. Mori, K. Storms, and S. C. Peters, "Conservative estimation of perception relevance of dynamic objects for safe trajectories in automotive scenarios," *ArXiv*, vol. abs/2307.10873, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259991178

[7] C.-H. Cheng, "Safety-aware hardening of 3d object detection neural network systems," in *International Conference on Computer Safety, Reliability, and Security*, 2020.

[8] O. Willers, S. Sudholt, S. Raafatnia, and S. Abrecht, "Safety concerns and mitigation approaches regarding the use of deep learning in safety-critical perception tasks," in *SAFECOMP Workshops*, 2020.

[9] A. D'Amour, K. A. Heller, D. I. Moldovan, B. Adlam, B. Alipanahi, A. Beutel, C. Chen, J. Deaton, J. Eisenstein, M. D. Hoffman, F. Hormozdiari, N. Houlsby, S. Hou, G. Jerfel, A. Karthikesalingam, M. Lucic, Y.-A. Ma, C. Y. McLean, D. Mincu, A. Mitani, A. Montanari, Z. Nado, V. Natarajan, C. Nielson, T. F. Osborne, R. Raman, K. Ramasamy, R. Sayres, J. Schrouff, M. G. Seneviratne, S. Sequeira, H. Suresh, V. Veitch, M. Vladymyrov, X. Wang, K. Webster, S. Yadlowsky, T. Yun, X. Zhai, and D. Sculley, "Underspecification presents challenges for credibility in modern machine learning," *J. Mach. Learn. Res.*, vol. 23, pp. 226:1–226:61, 2020.

[10] M. Wolf, L. R. Douat, and M. Erz, "Safety-aware metric for people detection," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 2759–2765.

[11] A. Bansal, J. Singh, M. Verucchi, M. Caccamo, and L. Sha, "Risk ranked recall: Collision safety metric for object detection systems in autonomous vehicles," in *2021 10th Mediterranean Conference on Embedded Computing (MECO)*. IEEE, 2021, pp. 1–4.

[12] M. Lyssenko, C. Gladisch, C. Heinzemann, M. Woehrle, and R. Triebel, "Towards Safety-Aware Pedestrian Detection in Autonomous Systems," in *Proc. of IROS*, Kyoto, Japan, Oct. 2022, pp. 293–300.

[13] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," *arXiv preprint arXiv:1903.11027*, 2019.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2999–3007.

[15] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. N. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digit. Signal Process.*, vol. 126, p. 103514, 2021.

[16] A. Ceccarelli and L. Montecchi, "Evaluating the consequences of object (mis)detection from a safety and reliability perspective: Discussion and measures," 2022.

[17] M. Lyssenko, C. Gladisch, C. Heinzemann, M. Woehrle, and R. Triebel, "From evaluation to verification: Towards task-oriented relevance metrics for pedestrian detection in safety-critical domains," in *Workshop on Safe Artificial Intelligence for Automated Driving*, 2021.

[18] S. S. Gannamaneni, S. Houben, and M. Akila, "Semantic concept testing in autonomous driving by extraction of object-level annotations from carla," *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1006–1014, 2021.

[19] S. Topan, K. Leung, Y. Chen, P. Tupekar, E. Schmerling, J. Nilsson, M. Cox, and M. Pavone, "Interaction-dynamics-aware perception zones for obstacle detection safety evaluation," in *2022 IEEE Intelligent Vehicles Symposium (IV)*, 2022, pp. 1201–1210.

[20] S. Topan, Y. Chen, E. Schmerling, K. Leung, J. Nilsson, M. Cox, and M. Pavone, "Refining obstacle perception safety zones via maneuver-based decomposition," in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pp. 1–8.

[21] Y. Tang, B. Li, M. Liu, B. Chen, Y. Wang, and W. Ouyang, "Autopedestrian: An automatic data augmentation and loss function search scheme for pedestrian detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 8483–8496, 2021.

[22] B. Li, Y. Yao, J. Tan, G. Zhang, F. Yu, J. Lu, and Y. Luo, "Equalized focal loss for dense long-tailed object detection," 2022.

[23] H.-C. Liao, C.-H. Cheng, H. Esen, and A. Knoll, "Improving the safety of 3d object detectors in autonomous driving using iogt and distance measures," 2023.

[24] S. Abrecht, A. Hirsch, S. Raafatnia, and M. Woehrle, "Deep learning safety concerns in automated driving perception," 2023.

[25] P. Schneider, M. Butz, C. Heinzemann, J. Oehlerking, and M. Woehrle, "Towards threat metric evaluation in complex urban scenarios," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 1192–1198.

[26] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, 2014.

[27] J. Dahl, G. R. de Campos, C. Olsson, and J. Fredriksson, "Collision avoidance: A literature review on threat-assessment techniques," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 1, pp. 101–113, March 2019.

[28] A. Badki, O. Gallo, J. Kautz, and P. Sen, "Binary TTC: A temporal geofence for autonomous navigation," *CoRR*, vol. abs/2101.04777, 2021.

[29] "nuscenes-devkit," https://github.com/nutonomy/nuscenes-devkit, 2021.

[30] "Possible issues with annotations," 2022, accessed on September 14, 2023. [Online]. Available: https://github.com/nutonomy/nuscenes-devkit/issues/366

[31] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Neural Information Processing Systems*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202786778

[32] Y. Henon, "pytorch-retinanet," https://github.com/yhenon/pytorch-retinanet, 2021.

[33] Z. Tian, "Fcos," https://github.com/tianzhi0549/FCOS, 2021.

[34] X. Zhou and O. Wu, "Which samples should be learned first: Easy or hard?" *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023.