Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 19/25

R. Chan, M. Rottmann, R. Dardashti, F. Hüger, P. Schlicht und H. Gottschalk

The Ethical Dilemma when (not) Setting up Cost-based Decision Rules in Semantic Segmentation

July 17, 2019

http://www.math.uni-wuppertal.de

The Ethical Dilemma when (not) Setting up Cost-based Decision Rules in Semantic Segmentation

Robin Chan[1], Matthias Rottmann[1], Radin Dardashti[2],
Fabian Hüger[3], Peter Schlicht[3] and Hanno Gottschalk[1]

[1]University of Wuppertal, School of Mathematics and Natural Sciences
[2]University of Wuppertal, Philosophical Seminar, Philosophy of Science
[3]Volkswagen Group Research, Automated Driving, Architecture and AI Technologies
{rchan,rottmann,dardashti,hgottsch}@uni-wuppertal.de
{peter.schlicht,fabian.hueger}@volkswagen.de

Abstract

Neural networks for semantic segmentation can be seen as statistical models that provide for each pixel of one image a probability distribution on predefined classes. The predicted class is then usually obtained by the maximum a-posteriori probability (MAP) which is known as Bayes rule in decision theory. From decision theory we also know that the Bayes rule is optimal regarding the simple symmetric cost function. Therefore, it weights each type of confusion between two different classes equally, e.g., given images of urban street scenes there is no distinction in the cost function if the network confuses a person with a street or a building with a tree. Intuitively, there might be confusions of classes that are more important to avoid than others. In this work, we want to raise awareness of the possibility of explicitly defining confusion costs and the associated ethical difficulties if it comes down to providing numbers. We define two cost functions from different extreme perspectives, an egoistic and an altruistic one, and show how safety relevant quantities like precision / recall and (segment-wise) false positive / negative rate change when interpolating between MAP, egoistic and altruistic decision rules.

1. Introduction

Machines acting autonomously in spaces co-populated by humans and robots are no longer a futuristic vision, but are part of the agenda of the world's technologically most advanced corporations. Autonomous car driving has seen spectacular advances due to recent progress in artificial intelligence (AI) and therefore is one of the corner-cases for this development. As street traffic, according to the world health organization (WHO), causes an annual death toll of 1.35M persons at the time of writing [20], it is expected that also autonomous driving cars will be involved in such tragic events. While there are reasons to believe that autonomous driving can reduce the overall numbers of deaths and heavy injuries, besides being required by *e.g.* the Ethics Commission instated by the German Federal Ministry of Transport and Digital Infrastructure [19], many further ethical issues remain in the choices of programming an autonomous vehicle. Therefore, autonomous cars have been a much-discussed topic in robot ethics [18], ranging from inevitable ethical dilemmas like the trolley problem [13, 16] to more mundane ethical situations [14].

In most of these ethical situations discussed in the literature, the robots and the AI algorithms controlling them are assumed to know the situation they decide on, whereas most deadly accidents with the involvement of self-driving cars in some way or another are connected with the (insufficient) perception of the vehicle's surrounding (see [4] for a preliminary report). Whether the AI algorithms of perception themselves depend on choices that involve ethical decisions is therefore a legitimate question.

For a practitioner in the field it is quite obvious that the answer is "yes": In semantic segmentation, the choice of training data, the selection of classes, potential class imbalance, the amount of data, the capacity of the learning algorithm and the performance of the hardware all determine what a contemporary AI algorithm is able to "see" and how error prone its perception will be. As errors in perception are potential root causes of accidents, ethical implications clearly exist.

In this work, we draw the attention to one further issue that is connected to the probabilistic output of semantic segmentation neural networks that are mostly used for the perceptive task. As the softmax output of a segmenta-

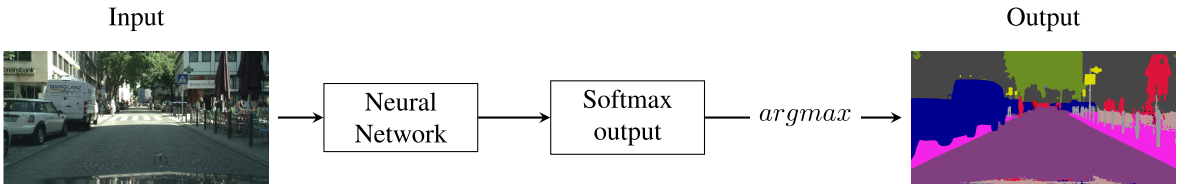

Figure 1. Illustration of semantic segmentation performed on an image of the Cityscapes dataset [10] with a neural network in combination with (pixel-wise) maximum a-posteriori probability classification.

tion network gives a pixel-wise class distribution, the maximum a-posteriori probability (MAP) principle, also known as Bayes decision rule, selects the class of highest probability. This is however not the only selection principle, as one could also apply the Maximum Likelihood (ML) decision rule that picks the class for which the input data is most representative [12]. While both rules have the appeal of being mathematically "natural", they are merely two examples of cost-based decision rules, where each confusion event is penalized by a specific quantity $c(\hat{k}, k)$ that valuates the aversion of a decision maker towards the confusion of the predicted class $\hat{k}$ with the actual class k. The decision on the predicted class now minimizes the expected cost.

Seen from this angle, the MAP principle corresponds to the cost matrix that attributes equal cost to any confusion event. We call this the robotistic valuation of the segmentation network's output. Human common sense would valuate the confusion of the street with a pedestrian differently from the confusion event with the roles interchanged: an unjustified emergency brake is a much weaker consequence as potential harm than overlooking a person on the street and therefore should come with a significantly lower cost. While it seems reasonable to assume that the confusion cost should be different from constant, it is ethically much less evident, which numbers should explicitly be used. In these situations of moral uncertainty, different ethical schools of thought may provide different answers, with some refusing to weigh lives at all [5]. In addition, legislation can put strong constraints on the choice as well. However, as the MAP principle and the ML decision rule already define confusion cost matrices, choices about these numbers have already been made. We, therefore, aim to make more transparent the ethical dimension involved in making a choice regarding a decision rule with its corresponding cost matrix.

We realize that the ultimate step from probabilities to perception depends on cost matrices in a high dimensional value space $\mathcal{V}$ and that the selected valuation changes the perception. Thereby, it also changes the consequences, as, *e.g.*, the precision and recall rates of specific classes. Furthermore, different cost matrices $C \in \mathcal{V}$ might express different ethical attitudes, like more egoistic (centred on the passenger in the (ego-) car) or altruistic (centred on public safety). Putting drivers first vs. putting the public first has already been subject to intense public debate [24].

In this paper, we do not intend to resolve the problem outlined above in any way. We present a numerical study that demonstrates the practical relevance of the problem by traveling through the value space within a triangle of robotistic and approximately egoistic and altruistic, respectively, cost value systems. Here the egoistic and altruistic cost matrices are set up in an *ad hoc* manner and are not meant to accurately represent these attitudes. Also, the matrices are by no means the most extreme ones spanning the value space. Nevertheless, when traveling through this small triangle in the large space of valuations $\mathcal{V}$, we see significant and relevant differences in the perception and measure consequences like the precision / recall and (segment-wise) false positive / negative rates for specific classes.

The remainder of this paper is organized as follows: In section 2 we describe our use-case for decision rules in neural networks, in particular in semantic segmentation neural networks. Next, in section 3 we explain the concept of decision rules in general and how they can be modified by valuating confusion costs between classes. We see various possibilities of defining the mentioned costs and provide two concrete examples in form of matrices in section 4. Moreover, we present our spanned value space of confusion cost matrices and the setup for our experiments which follow in section 5. We show that different cost matrices are capable of considerably affecting the perception of a state-of-the-art semantic segmentation network in the setting of urban street scenes.

2. Standard decision rule in neural networks

Semantic segmentation is the task of assigning each pixel of an image to one of the predefined classes $\mathcal{K} = \{1, \ldots, N\}$. Suppose, we use a neural network for solving this task. Let $x \in \{(r, g, b)\}^{m \times n}$, $(r, g, b) \in \{0, \ldots, 255\}^3$ be an "rgb" (red, green and blue light additively colored) input image with resolution $m \times n$. After processing the image x with a neural network we obtain a posterior probability distribution $p_{ij}(k|x)$ over all classes $k \in \{1, \ldots N\}$ at location (pixel position in the image) $(i, j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}$. The 3D tensor $p_{ij}(k|x)$ represents the softmax output of a neural network for semantic segmentation. The third dimension is given by the choice of $k \in \{1, \dots N\}$. This provided probability distribution expresses the confidence of the neural network as statistical prediction model to label the input correctly given the class k. The pixel-wise classification is then performed by applying the $argmax$ function (pixel-wise) on the posterior probabilities / softmax output. This kind of decision making is called maximum a-posteriori probability (MAP) principle.

In the field of Deep Learning, following the MAP as *decision rule* is by far the most commonly used one. It maximizes the overall performance of a neural network, meaning in cases of large prediction uncertainty, this rule tends to predict classes that appear frequently in the dataset. However, classes of potential high importance, like in autonomous driving the classes traffic signs and humans, usually appear less frequently. These classes are rare in terms of the number of instances and the number of pixels in the dataset. This problem is in close connection to the fact that the MAP estimation considers all prediction mistakes to be equally serious which is in conflict with human intuition. Thus, a natural approach is to weight different prediction mistakes against each other.

3. Cost-based decision rules in neural networks

Let Ω be a population consisting of $N \geq 2$ disjoint subsets. For each element $\omega \in \Omega$ we assume there exists one feature vector $x(\omega) \in S \subset \mathbb{R}^n$. Let

$$X : \Omega \to S \tag{1}$$

$$K : \Omega \to \{1, \dots, N\} = \mathcal{K} \tag{2}$$

be random variables for feature vector x and class affiliation k, respectively. A *decision rule* can be defined as a map

$$d : \quad S \to \mathcal{K} \tag{3}$$

$$x(\omega) \mapsto \hat{k}(\omega) \tag{4}$$

which assigns an element from the feature space to one class. We say, $d(x) = \hat{k}$ is the predicted class for feature vector x. Furthermore, we describe the a-posteriori probability of an object to belong to class k given feature x as

$$p(k|x) := P(K = k \mid X = x). \tag{5}$$

Usually, this probability is not known and needs to be estimated. We assume in the following that this is already accomplished, *e.g.*, $p(k|x)$ is approximated by the softmax output of a neural network.

Cost-based decision rules follow the idea of assigning one input to the class which minimizes the expected cost given one confusion cost function

$$c : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_{\geq 0} := [\, 0, \infty \,).$$

Considering all possible confusion cases we obtain a confusion cost matrix

$$C := (c(\hat{k}, k))_{\hat{k}, k = 1, \dots, N} \in \mathcal{V} \subset \mathbb{R}^{N \times N}_{\geq 0} \tag{6}$$

with $\hat{k}$ being the predicted class while k being the target class and

$$\mathcal{V} := \{\, C \in \mathbb{R}^{N \times N} \mid C_{jj} = 0, C_{ij} > 0, i, j \in \mathcal{K} \,\} \tag{7}$$

being the value space of all valid matrices C for cost-based decision rules. Hence, all elements of a valid matrix must be positive except the diagonal elements, which must equal 0, according to $\mathcal{V}$. Strictly speaking, $\mathcal{V}$ consists of equivalence classes since each C in combination with cost-based decision rules will produce the same output as $\mu C, \mu > 0$, *i.e.*, different scales of C do not change the output. Therefore, rather the costs of the classes relative to each other are decisive for the output instead of the absolute values.

In order to understand the just stated fact we define the expected cost with respect to confusion cost functions via

$$\mathbb{E}[\, c(k', K) \mid X = x \,] = \sum_{k=1}^{N} c(k', k)\, p(k|x) \tag{8}$$

and the corresponding *cost-based* decision rule as

$$d(x; C) := \underset{k' \in \{1, \dots, N\}}{\operatorname{argmin}} \sum_{k=1}^{N} c(k', k)\, p(k|x) \tag{9}$$

$$\overset{(6)}{=} \underset{k' \in \{1, \dots, N\}}{\operatorname{argmin}} \; C_{k'} \cdot \vec{p}(x) = \hat{k} \tag{10}$$

with $C_k := (C_{k1}, \dots, C_{kN})$ being the k-th row vector of $C \in \mathcal{V}$ and $\vec{p}(x) := (p(1|x), \dots, p(N|x))^T$ being the posterior probabilities vector conditioned on the feature x. This rule is optimal considering the expected costs.

Cost-based decision rules are strongly related to probability thresholding. The aim of probability thresholding is to make class predictions cost-sensitive during inference by moving the output threshold towards inexpensive classes. This is achieved by defining a confusion cost function of the form

$$c\left(\hat{k}, k\right) := \begin{cases} 0 & , \text{ if } \quad \hat{k} = k \\ \psi(k) & , \text{ if } \quad \hat{k} \neq k \end{cases} , \; \psi(k) \in \mathbb{R}_{\geq 0} \tag{11}$$

with $\psi(k) > \psi(k')$ if we want the network to prefer predicting class k to predicting class k'. One special type of c is the simple symmetric cost function [12]

$$c_s\left(\hat{k}, k\right) := \begin{cases} 0 & , \text{ if } \quad \hat{k} = k \\ \lambda & , \text{ if } \quad \hat{k} \neq k \end{cases} , \; \lambda \in \mathbb{R}_{\geq 0} \tag{12}$$

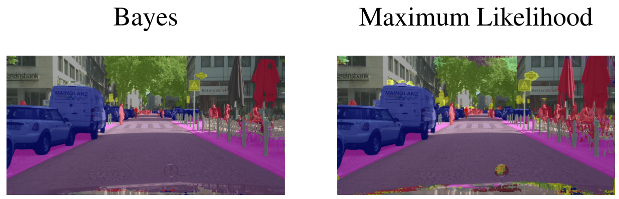

Figure 2. Illustration of two segmentation masks obtained with the Bayes decision rule (right) and the Maximum Likelihood decision rule (left). The difference between these two masks lies in the adjustment with the (pixel-wise) prior class probabilities in the decision rule during inference.

whose incorporation in the cost-based decision rule is equivalent to the MAP principle. Given $c_s(\hat{k},k)$ all elements in the confusion cost matrix C_s are equal to the constant λ except the diagonal elements which are equal 0. Accordingly, the cost-based decision rule takes the form:

$$d(x;C_s) \overset{(12)}{=} \operatorname*{argmin}_{k\in\{1,\dots,N\}} \sum_{k'=1,k'\neq k}^{N} \lambda\cdot p(k'|x) \tag{13}$$

$$= \operatorname*{argmin}_{k\in\{1,\dots,N\}} \; 1-p(k|x) \tag{14}$$

$$= \operatorname*{argmax}_{k\in\{1,\dots,N\}} \; p(k|x) \; =: \; d_{Bayes}(x). \tag{15}$$

In decision theory equation (15) is the definition of the *Bayes* decision rule which is equivalent to the MAP principle and therefore also to the default classification principle in neural networks. However, the simple symmetric cost function implies an equal class weighting, *i.e.*, weighting every confusion between two classes (or each type of misclassification) equally. Depending on the purpose, this setting does not reflect the intuition of most people but is still applied in most deep learning state-of-the-art models.

A mathematically natural way to approach this problem is exchanging the simple symmetric with the inverse proportional cost function [12] which is another special type of c. In light of confusion costs the latter cost function

$$c_p\left(\hat{k},k\right) := \begin{cases} 0 & \text{, if} \quad \hat{k}=k \\ \lambda/p(k) & \text{, if} \quad \hat{k}\neq k \end{cases}, \; \lambda\in\mathbb{R}_{\geq 0} \tag{16}$$

weights each confusion with the inverse prior probability $1/p(k), p(k)\in(0,1)$ of the potential target class k. In neural networks the class appearance frequencies in the training data correspond approximately to the priors. Considering the priors, we can put more emphasis on finding classes which are rare, *i.e.*, classes which have a low prior probability. The decision rule resulting from this is the *Maximum Likelihood* (ML) decision rule

$$d_{ML}(x) := \operatorname*{argmax}_{k\in\{1,\dots,N\}} \; p(x|k). \tag{17}$$

road	road
flat	sidewalk terrain
static	building wall fence pole motorcycle bicycle
info	traffic light traffic sign
humans	person rider
dynamic	car truck bus train

Figure 3. Class aggregates of Cityscapes classes that we use for simplicity in our experiments. Note that in the Cityscapes labeling motorcycles and bicycles in motion adhere to the class “rider”.

Now x is mapped to the class k for which the observed features are most typical, independent of a prior belief about the class frequencies. As presented in [6], with respect to rare classes the application of the ML rule significantly reduces the number of false negative (overlooked) segments for rare classes, but to the detriment of producing substantially more false positive segment predictions. One might argue that there is a “sweet spot” where the two error rates, the positive and negative one, are optimal. However, one might also argue that certain classes are still underweighted relative to others. We address both problems by applying the cost-based decision rule in combination with adjusting the confusion cost matrix C.

4. Setup of experiments

For our experiments we use the Cityscapes dataset with 19 semantic classes. In order to reduce the number of confusion cost values to be specified for the matrix C we aggregate classes that are treated similarly considering confusion costs, see figure 3 for a first attempt although refined aggregations are probably more appropriate.

With 6 aggregated classes we define a 6×6 matrix. For performance evaluation we map the reduced matrix back to full 19×19 size such that all combinations between classes out of two aggregates have an equal confusion cost, *i.e.*, for two different non-empty aggregates $\mathcal{I},\mathcal{J}\subset\mathcal{K}$ it holds

$$\mathcal{I}\cap\mathcal{J}=\emptyset \tag{18}$$

$$\Leftrightarrow c(i,j)=c(i',j')\;\forall\, i,i'\in\mathcal{I},\; j,j'\in\mathcal{J}. \tag{19}$$

In addition, we set a small $\epsilon = 0.1$ for all confusions between different classes within an aggregate so that we apply the Bayes decision rule (only within an aggregate) without affecting the cost-based decision between aggregated classes, *i.e.*, for each non-empty aggregate $\mathcal{I}\in\mathcal{K}$ it holds

$$c(i,i')=\epsilon\;\forall\, i\neq i'\in\mathcal{I} \tag{20}$$

$$c(i,i)=0\;\forall\, i\in\mathcal{I}. \tag{21}$$

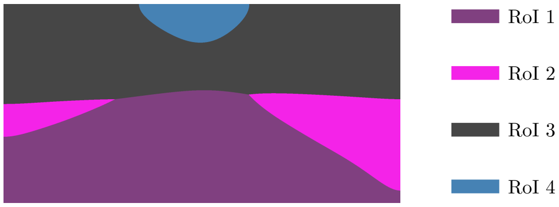

Figure 4. Regions of interest derived from the priors of the classes building, road, sidewalk and sky in the Cityscapes dataset.

Note that we suppress the "sky" class in our class aggregation although it is one of the originally trained classes. The reason is that we believe that overlooking the sky does not result in dangerous traffic scenarios. Therefore, we prevent the network from predicting sky by setting $C^T_{sky} = \{M\}^N$ with $M = 1000$ being a sufficiently large cost value. This implies that the confusion of any (target) class with sky is valuated with high cost. We set the cost for the converse confusion, when sky is the target class, to a constant value in order to not affect the class prediction between the remaining classes.

To gain further insight we define image regions of interest (RoI). These regions are derived from the pixel-wise class frequencies (priors) of the classes "road", "sidewalk", "building" and "sky" in the Cityscapes dataset. We obtain the 4 regions of interest (or 5 regions as the sidewalk RoI consists of two connected components) by assigning each pixel to the class with the highest class appearance frequency at the corresponding pixel location, see figure 4.

For our experiments we further define two confusion cost matrices representing two extreme views in traffic scenes. On the one hand, we define the "altruistic" matrix C_A that prioritizes all traffic participants and particularly humans. On the other hand, we define the "egoistic" matrix C_E that only prioritizes the safety and comfort of the passenger inside the (ego-) car. The chosen cost values can be viewed in figure 6. We compare the corresponding predictions with each other and also with the Bayes rule's prediction, respectively. The Bayes decision rule implies the matrix $C_R := (c_s(\hat{k}, k))_{\hat{k},k=1,\dots,N}$ which we term in the following the "robotistic" confusion cost matrix. This method is robotistic in the sense that, in any event, the only goal is to minimize all error rates. The convex combinations of these three presented matrices span a confusion value space

$$\begin{aligned} V := \{\, C \in \mathcal{V} \mid \alpha C_R + \beta C_A + \gamma C_E = C, \\ \alpha + \beta + \gamma = 1,\ \alpha, \beta, \gamma \geq 0 \,\} \end{aligned} \tag{22}$$

(see figure 7 and figure 8). It is important to emphasize that $V \subset \mathcal{V}$ is only one subspace of a far bigger possible value space. There are even more extreme cost matrices that enlarge the space dramatically. There are also cost matrices expressing views in a completely different direction and therefore increasing the dimensionality of the space. However, our presented V is sufficient in order to show that it is already capable of changing our model's perception significantly.

Cost matrix	Class	RoI	Precision	Recall
Altruistic	Person	1	41.12%	**99.81**%
Robotistic	Person	1	89.87%	94.98%
Egoistic	Person	1	**93.88**%	70.07%
Altruistic	Person	2	39.42%	**99.86**%
Robotistic	Person	2	88.36%	93.93%
Egoistic	Person	2	**95.07**%	54.81%
Altruistic	Building	1	22.56%	93.65%
Robotistic	Building	1	**80.99**%	94.94%
Egoistic	Building	1	15.15%	**99.93**%
Altruistic	Building	2	24.94%	95.22%
Robotistic	Building	2	**87.76**%	94.58%
Egoistic	Building	2	18.48%	**99.90**%

Table 1. Precision and recall rates for the three different cost matrices. The rates are computed for the classes person and building in the street and the sidewalk RoIs, *i.e.*, RoI 1 and 2.

5. Experiments

As part of autonomous car driving systems, interpreting visual inputs is crucial in order to obtain a full understanding of the car's environment. The inference of an image in semantic segmentation [10, 11] is performed at pixel level combining object detection and localization. In recent years, deep learning has achieved great success in a wide range of problems including semantic segmentation. Most state-of-the-art models are built on deep convolutional neural networks (CNNs) [15, 23]. One important contribution to CNNs for semantic segmentation is the Fully Convolutional Network (FCN) [22] which introduces end-to-end training taking input of arbitrary size and producing output of equal size. The network is one of the first using an encoder-decoder structure [3, 21] whose encoder part is a classification network followed by the decoder part that projects convolved learned features back onto full pixel resolution. With the integration of atrous (also called dilated) convolutions [25], that allows an exponential increase of the network's receptive field without loss of resolution, the performance of semantic segmentation networks is further significantly improved. One advanced module based on the latter operation is atrous spatial pyramid pooling (ASPP) [7]. It is one of the main contributions to the network DeepLabv3+ [8] which we use in the following in our experiments.

We demonstrate the performance of cost-based decision rules with different confusion cost matrices on the Cityscapes [10] validation dataset. DeepLabv3+ is already pretrained on the latter dataset and implemented in TensorFlow [1]. The implementation and tuned weights are pub-

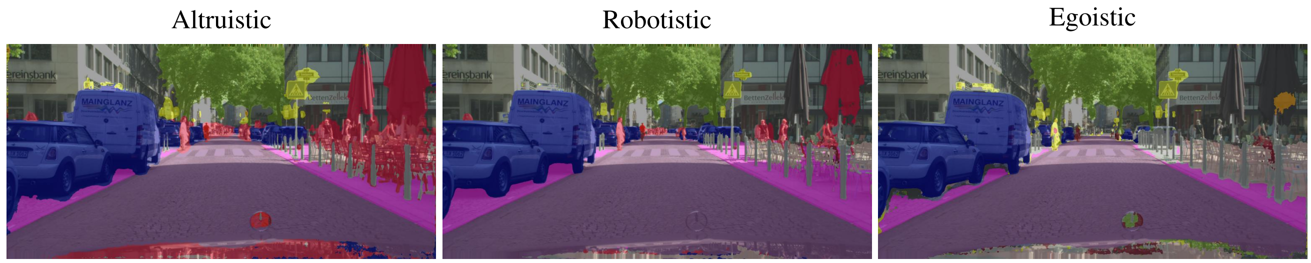

Figure 5. Illustration of three semantic segmentation masks and different perception obtained by the application of cost-based decision rules with an altruistic, a simple symmetric (robotistic) and an egoistic cost matrix.

$$C_A = \begin{array}{cccccccc} & \text{“road”} & \text{“flat”} & \text{“static”} & \text{“info”} & \text{“human”} & \text{“dynamic”} & \\ \left(\begin{array}{c}\\ \\ \\ \\ \\ \\ \end{array}\right. & & & & & & & \end{array}$$

$$C_A = \left(\begin{array}{cccccc} 0 & 10^0 & 10^1 & 10^2 & 10^3 & 10^2 \\ 10^0 & 0 & 10^1 & 10^2 & 10^3 & 10^2 \\ 10^0 & 10^0 & 0 & 10^2 & 10^2 & 10^1 \\ 10^0 & 10^0 & 10^0 & 0 & 10^3 & 10^2 \\ 10^0 & 10^0 & 10^0 & 10^2 & 0 & 10^1 \\ 10^0 & 10^0 & 10^0 & 10^2 & 10^3 & 0 \end{array}\right) \begin{array}{l} \text{“road”} \\ \text{“flat”} \\ \text{“static”} \\ \text{“info”} \\ \text{“human”} \\ \text{“dynamic”} \end{array}$$

$$C_E = \overbrace{\left(\begin{array}{cccccc} 0 & 10^0 & 10^3 & 10^2 & 10^1 & 10^2 \\ 10^0 & 0 & 10^3 & 10^2 & 10^1 & 10^2 \\ 10^1 & 10^0 & 0 & 10^3 & 10^0 & 10^1 \\ 10^1 & 10^0 & 10^3 & 0 & 10^0 & 10^1 \\ 10^1 & 10^0 & 10^3 & 10^2 & 0 & 10^2 \\ 10^1 & 10^1 & 10^3 & 10^2 & 10^2 & 0 \end{array}\right)}^{\textit{potential target in columns}} \left.\begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array}\right\} \textit{prediction in rows}$$

Figure 6. Two extreme confusion cost matrices that we study in our experiments. C_A represents the altruistic view prioritizing all traffic participants and particularly pedestrians. C_E represents the egoistic view prioritizing only the passenger in the (ego-) car. One element in the matrix expresses the cost that arises if we predict the class corresponding to the row and we confuse it with the potential target class corresponding to the column.

licly available on GitHub. As network backbone, we choose the modified version of the Xception model [9] that attains an mIoU score of 79.55% on the Cityscapes validation set with the application of the MAP / Bayes decision rule.

In the following, we perform our analysis for the classes “person” and “building” which are key classes in our problem setting of autonomous driving for the altruistic and egoistic view, respectively. Furthermore, we focus our studies on the regions of interest 1 & 2, the near field perception in front of the (ego-) car and to the side of the (ego-) car.

Pixel-wise precision vs. recall. For evaluation we first consider precision and recall. These two metrics are closely connected to the quantities false positive and false negative pixel predictions. A predicted pixel is a false positive (FP) if it falsely indicates an object’s presence. A predicted pixel ignoring the presence of a present object is a false negative (FN). Therefore, precision is the percentage of a model’s predicted pixels that match the ground truth, while recall is the percentage of ground truth pixels that a model predicts correctly, *i.e.*,

$$prc = TP \,/\, (TP + FP) \tag{23}$$
$$rec = TP \,/\, (TP + FN) \tag{24}$$

with TP being the true positives (pixels correctly classified according to the ground truth). The two evaluation metrics can be formulated as maps

$$prc, rec : V \to [\,0, 1\,] \tag{25}$$

expressing the neural network’s predictive power depending on $C \in V$. The higher the value, the less prediction mistakes we obtain regarding falsely detected and non-detected pixels, respectively. The precision and recall scores of the different cost matrices in different regions of interest can be found in table 1.

For the class person we observe that the recall is maximized when using C_A. Compared to C_R the reduction is 4.83 percent points in the street RoI and even 5.93 percent points in the sidewalk RoI. Even if the recall of person instances is already impressively high, C_A is still capable of boosting the performance in this metric such that nearly no person pixels are missed. However, to a striking detriment, the precision is reduced by about 48 percent points in both RoIs down to 41.12% and 39.42%, respectively. When using C_E persons are ignored to a large extent leading to a recall reduction of 24.91 percent points in the frontal RoI and 39.12 percent points in the sidewalk RoI in comparison to C_R. Consequently, the precision is increased by 4.01 and 6.71 percent points, respectively. With C_E DeepLabv3+ only predicts persons if the network indicates a high confidence about its decision. As expected there is a trade off between the metrics, *i.e.*, increasing one performance measure decreases the other and vice versa. Also noteworthy

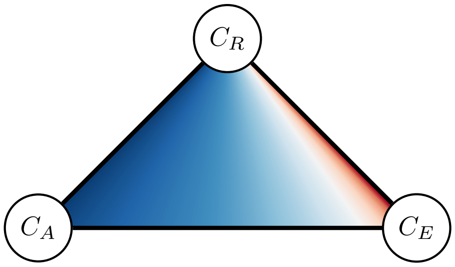

Figure 7. Confusion cost matrix space V spanned by our exemplary altruistic (C_A) and egoistic (C_E) cost matrix and the robotistic (C_R) cost matrix. Inside the triangle as heatmap the behavior of *rec*($V(C)$ | *person*), the recall of person pixels. Blue indicates high recall, red indicates low recall.

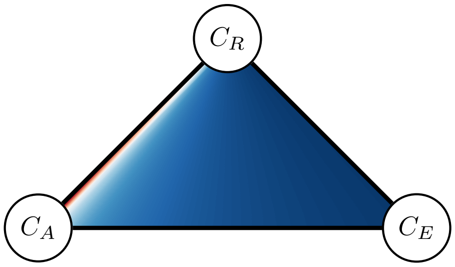

Figure 8. Confusion cost matrix space V spanned by our exemplary altruistic (C_A) and egoistic (C_E) cost matrix and the robotistic (C_R) cost matrix. Inside the triangle as heatmap the behavior of *rec*($V(C)$ | *building*), the recall of building pixels. Blue indicates high recall, red indicates low recall.

from this analysis is that DeepLabv3+ confuses only persons which are not completely visible, *e.g.*, persons standing behind cars or around corners. Only small parts of person instances are mainly overlooked, see also figure 9.

For the class building we also observe this trade off but only between C_E and C_R. C_E improves the recall by 4.99 and 5.32 percent points while reducing the precision substantially by 65.84 and 69.28 percent points, respectively, for the street and the sidewalk RoI.

The behavior is different with respect to C_A. Regarding building segments, C_R performs better in both metrics in the frontal RoI. The recall is reduced by 1.29 and the precision by significant $58,43$ percent points. In the sidewalk RoI, the recall of C_A is slightly improved (0.64%) but the precision is again drastically reduced to 24.94%. Noteworthy from this analysis is that DeepLabv3+ has difficulties in detecting separated ground truth segments of building instances which arise from objects in front of buildings and splitting the instance's actual connected component in the ground truth, see also figure 10.

Segment-wise false-detection vs. non-detection. Another interesting quantity are the entire false-detections and non-detections of person and building segments when using the different cost matrices. In this regard, we now define a segment to be, depending on the considered prediction or ground truth mask, a false positive / negative if the segment's intersection over union (*IoU*) equals 0. Figure 9 and figure 10 visualize the segments with $IoU = 0$ in the prediction mask and ground truth mask, respectively, again for the classes person and building. The presented heatmaps visibly confirm the findings from the precision and recall analysis. The application of cost-based decision rules changes the perception of DeepLabv3+ significantly. For instance, for the class person the altruistic cost matrix overproduces false positives but there are almost no overlooked person segments. On the contrary, the egoistic cost matrix almost completely refuses to predict the class person but is mostly correct in case it predicts a person segment. The robotistic cost matrix offers a balanced compromise between both prediction mistakes. Depending on people's individual sense of how the cost matrix should be defined, the presented observations will change again. Thus, what will remain open is a concrete suggestion to the inevitable definition of a confusion cost matrix.

6. Discussion

In this paper we illustrated the impact of cost-based decision rules on the perception of a state-of-the-art semantic segmentation neural network. In this framework, we discussed options for setting up cost-based decision rules ranging from the classical "robotistic" maximum a-posteriori probability principle over probability thresholding and the Maximum Likelihood decision rule to *ad hoc* "egoistic" and "altruistic" cost assignments to confusion events. Within the triangle of robotistic, egoistic and altruistic attitudes, we investigated precision and recall and also false positive and negative rates in two regions of interest for the classes "person" and "building" in the Cityscapes dataset. We demonstrated the metrics' dependence on the convex combination of the cost matrices from the three mentioned ethical attitudes spanning a triangle within a larger space of values.

On the technical side, many questions concerning the use of cost-based decision rules have to be clarified, *e.g.* the adaptation of cost matrices to prior probabilities or the impact on "downstream" modules like data fusion with other sensors and trajectory planning.

Let us turn to the ethical side of the discussion. The probabilistic nature of the output of the segmentation network makes a decision rule necessary. As different decision rules have non-converging consequences, a choice for a decision rule amounts to a choice where in the long run human lives are weighted against other considerations. This choice is therefore not one to be made from a purely technical side (by *e.g.* choosing the mathematically "natural" decision rule) but one that needs to recognize its ethical di-

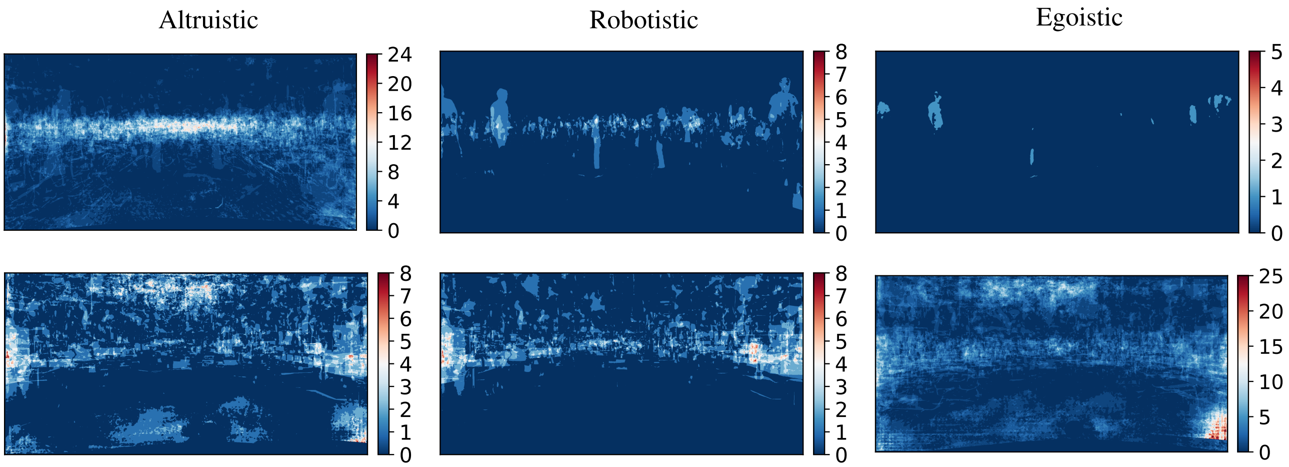

Figure 9. Falsely detected (false positive) person (top row) and building (bottom row) segments.

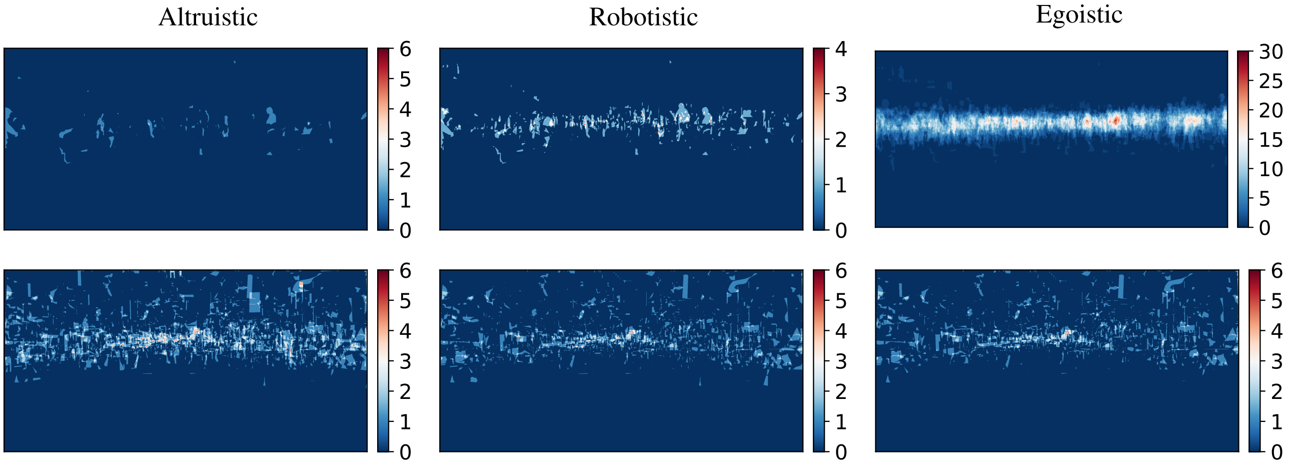

Figure 10. Non-detected (false negative) person (top row) and building (bottom row) segments.

mension. While technological advances may have an impact on these considerations they will not make the need for a decision rule obsolete.

This leads to the question: Which decision rule is the "right" one? As in most cases of moral uncertainty, different normative ethical schools of thought will provide different answers (see [17, Ch.3] for a short non-technical introduction in the context of robot ethics). A deontological strategy would try to justify a certain choice of a decision rule by arguing for the rule itself being ethically "good", not considering what may follow from that choice. For instance, a strict rule-based implementation of the requirement by the ethics commission that "[t]he protection of individuals takes precedence over all other utilitarian considerations." [19] may be interpreted to lead to a cost function that is never allowed to confuse a human for another object. A consequentialist strategy justifies a cost function by focusing on the consequences of a certain choice. This would involve the above analysis of the consequences of the egoistic and altruistic cost functions. Another approach refers to polling, using the ethical intuition of the majority of the people being asked. This can lead to strong cultural differences, as resulted in an analysis of Awad et al. in the context of trolley-like problems [2].

It is not the aim of this paper to defend any specific approach or to provide an alternative answer to the above problem of choosing the "right" decision rule, but to make transparent the underlying ethical dimension of what may seem as mathematically innocuous "natural' choices. This transparency is a precondition for a responsible handling and open debate on these issues.

Acknowledgment. R. Chan, M. Rottmann and H. Gottschalk acknowledge (partial) funding by Volkswagen Group Research through the contract "Maximum likelihood and cost-based decision rules in semantic segmentation".

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 5

[2] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, and I. Rahwan. The moral machine experiment. *Nature*, 563:59–64, 2018. 8

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015. 5

[4] N. T. S. Board. Preliminary report highway hwy18mh010, 2018. 1

[5] J. Broome. *Weighing lives*. Oxford University Press, 2004. 2

[6] R. Chan, M. Rottmann, F. Hüger, P. Schlicht, and H. Gottschalk. Application of decision rules for handling class imbalance in semantic segmentation. *CoRR*, abs/1901.08394, 2019. 4

[7] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *CoRR*, abs/1606.00915, 2016. 5

[8] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611, 2018. 5

[9] F. Chollet. Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357, 2016. 6

[10] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2, 5

[11] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, Jan 2015. 5

[12] L. Fahrmeir, A. Hamerle, and W. Häussler. *Multivariate statistical Methods (in German)*. Walter De Gruyter, 2 edition, 1996. 2, 3, 4

[13] P. Foot. The problem of abortion and the doctrine of double effect. *Oxford Review*, 5:5–15, 1967. 1

[14] J. Himmelreich. Never mind the trolley: The ethics of autonomous vehicles in mundane situations. *Ethical Theory and Moral Practice*, 21(3):669–684, 2018. 1

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. 5

[16] P. Lin. Why ethics matters for autonomous cars. In *Autonomous driving*, pages 69–85. Springer, Berlin, Heidelberg, 2016. 1

[17] P. Lin, K. Abney, and G. A. Bekey. *Robot ethics: the ethical and social implications of robotics*. The MIT Press, 2014. 8

[18] P. Lin, K. Abney, and R. Jenkins. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press, 2017. 1

[19] E. C. on automated, networked driving of the German Federal Ministry for Transport, and Infrastructure. Report of the ethics commission automated and networked driving (in german), 2017. 1, 8

[20] W. H. Organization. Road traffic injuries, 2018. 1

[21] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. 5

[22] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *PAMI*, 2016. 5

[23] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. 5

[24] M. Taylor. Self-driving mercedes-benzes will prioritize occupant safety over pedestrians. *Car and Driver*, Oct. 7, 2016. 2

[25] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *CoRR*, abs/1511.07122, 2015. 5