InfoSalGAIL: Visual Attention-Empowered Imitation Learning of Pedestrian Behavior in Critical Traffic Scenarios

Igor Vozniak, Matthias Klusch, André Antakli, Christian Müller
German Research Center for Artificial Intelligence (DFKI),
Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany
{*firstname.lastname*}*@dfki.de*

Keywords: Visual Attention-Empowered Imitation Learning, end-to-end human-like data-driven simulation, critical scenario generation.

Abstract: The imitation learning of complex pedestrian behavior based on visual input is a challenge due to the underlying large state space and variations. In this paper, we present a novel visual attention-based imitation learning framework, named InfoSalGAIL, for end-to-end imitation learning of (safe, unsafe) pedestrian navigation policies through visual expert demonstrations empowered by eye fixation sequence and augmented reward function. This work shows the relation in latent space between the policy estimated trajectories and visual-attention map. Moreover, the conducted experiments revealed that InfoSalGAIL can significantly outperform the state-of-the-art baseline InfoGAIL. In fact, its visual attention-empowered imitation learning tends to much better generalize the overall policy of pedestrian behavior leveraging apprenticeship learning to generate more human-like pedestrian trajectories in virtual traffic scenes with the open source driving simulator OpenDS. InfoSalGAIL can be utilized in the process of generating and validating critical scenarios for adaptive driving assistance systems.

1 INTRODUCTION

Complex human-like on-street walking activities are hard to mimic by means of end-to-end imitation learning. The underlying massive state-space and high variation in data may lead to insufficient objective generalization of the trained policy. Nevertheless, imitation learning based methods succeeded in a wide range of problems (Ziebart et al., 2008; Englert and Toussaint, 2018; Finn et al., 2016; Stadie et al., 2017; Ermon et al., 2015) given expert demonstrations. For example, the recently proposed system InfoGAIL (Li et al., 2017) performs end-to-end imitation learning based on clustered visual demonstrations by various experts. During training, the derivation of latent code relies on highly mutual information between the code and those expert demonstrations used for policy inference. However, our experiments revealed that InfoGAIL suffers from poor policy generalization in due course of learning to imitate human-like on-street walking pedestrians in complex traffic scenes.

To this end, we developed a novel approach for visual attention-empowered end-to-end imitation learning of pedestrians in virtual traffic scenes. The resulting system, named InfoSalGAIL, is now capable of learning given human expert behavior of on-street walking in complex traffic scenes based on visual input only with additional prior knowledge in the form of visual attention or saliency maps (cf. Section 3.2). Our experiments with human pedestrian behavior learning avatars in traffic scenes with the virtual driving simulator OpenDS[1] revealed that InfoSalGAIL may significantly fasten and improve the imitation learning process as compared to its baseline InfoGAIL (cf. Section 4.1). The code for reproducing the experiments is available in free access as part of the open source simulation software OpenDS (version 6.0). The contributions of this work are threefold:

1. Development of a novel approach to visual attention-empowered imitation learning based on integrated use of individual pedestrian saliency maps (cf. Figure 1) for large state spaces in complex urban traffic scenes. Experimental evaluation of the approach shows significant improvement over its baseline InfoGAIL in terms of speed and quality of learning.
2. Development of a new open-source benchmark for training and testing pedestrian imitation learn-

[1] https://opends.dfki.de

ing avatars in critical traffic scenarios with the open-source virtual driving simulator OpenDS. For this purpose, OpenDS has also been extended with new modules for covering human pedestrian-centred perspectives and control such as head movements (pitch, yaw) and walking.

3. The results of our experimental performance evaluation of the approach based on the German In-depth Accident Study (GIDAS[2]) strongly supports the hypothesis that the used saliency maps can be considered as a kind of individual pedestrian movement policy fingerprints. Their integrated usage may enforce the policy generator of pedestrian avatars to more human-like (expert-like) actions in virtual critical traffic scenarios.

Figure 1: Visual attention-empowered imitation learning with InfoSalGAIL in OpenDS traffic scene. The augmented eye fixation sequence in the center of the image shows those features that are most attractive for the individual pedestrian walking in the scene.

The remainder of the paper is structured as follows. Section 2 reviews the background required to follow this research work and covers the state-of-the-art analysis in the domain of imitation learning. Section 3 presents InfoSalGAIL, our visual attention-empowered imitation learning framework, which is applied to the problem of human-like pedestrian trajectory generation. Moreover, it covers the training and the testing phases with the explanation of the introduced loss function. Section 4 describes the data creation process as well as conducted experiments and obtained results in comparison to the chosen baseline. Finally, we conclude the paper in Section 5.

[2]https://www.gidas.org

2 BACKGROUND AND RELATED WORK

2.1 Visual Attention

According to Corbetta (1998), visual attention defines the mental ability to select stimuli, responses, memories or thoughts that are behaviorally relevant among the many others that are behaviorally irrelevant. The human visual system is agitated by entities of the surroundings in color and shape. It leads to the physiological fact that the human is not capable of attending at all elements of the visual field of view (FoV) at the same time due to the limited cognitive capacity of the brain. Thus, the brain has an ability to filter out visually perceived information, which was defined by Sully (1891) as selective visual attention, where the spotlight model has been initially proposed to justify the visual attention feature of the brain.

There are two classes of factors which influence visual attention, namely bottom-up and top-down. The bottom-up class is based on the physical properties of the objects that fall in the visual FoV of humans like shapes, colors, size and orientation. In contrast, the top-down class is task-dependent and influenced by the current task, cognitive abilities and/or experience. The prominent computational model of visual attention that describes it from both the bottom-up and top-down perspectives is the so-called saliency model. Its task is to identify an area of interest (most attractive or probable area), which can be seen as a set of pixels, which corresponds to the scene entities. We adopted a heat map (2D image of the same size), where the intensity of the color corresponds to the relevance of the given pixel. Figure 2 shows a sample of the expert‘s FoV and the corresponding saliency map.

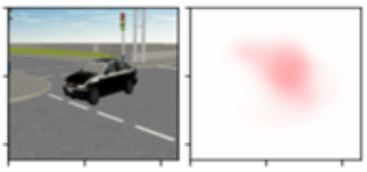

Figure 2: Visual attention sample. Left: expert‘s field of view of size 224×224 pixels; Right: corresponding saliency map. As seen from the given pair of images, a significant portion of visual attention map, given crossing the street task, is attached to the approaching vehicle.

2.2 Preliminaries

The pedestrian navigation problem, from the perspective on imitation learning (cf. Section 2.3), can

be defined as an infinite-horizon discounted Markov Decision Process (MDP) (S,A,P,r,ρ_0,γ), where S denotes the finite state space, A the finite set of actions, $P : S \times A \times S \Rightarrow \mathbb{R}$ the transition probability distribution function, $r : S \Rightarrow \mathbb{R}$ the reward function, $\rho_0 : S \Rightarrow \mathbb{R}$ the distribution of initial state s_0, and $\gamma \in (0,1)$ the discount factor. Let π denote a stochastic policy $\pi : S \times A \Rightarrow [0, 1]$ with the expert policy π_E to be mimicked given in the form of visual demonstrations only. A set of expert trajectories τ_E generated by the policy π_E consists of a sequence of state-action pairs. The expectation with respect to the policy π is used to denote an expectation with respect to the generated trajectories: $\mathbb{E}_\pi[f(s,a)] \triangleq \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t f(s_t,a_t)]$, such that $s_0 \sim \rho_0$, $a_t \sim \pi(a_t|s_t)$, $s_{t+1} \sim P(s_{t+1}|a_t,s_t)$.

2.3 Imitation Learning

In this work, we apply imitation learning to let pedestrian avatars learn to best mimic some demonstrated human pedestrian walking or navigation policy $\pi(a|s)$ without knowing the reward function r of the considered MDP. The two broadly applicable classes of solutions of imitation learning approaches are behavior cloning (BC) and apprenticeship learning (AL). Behavior cloning uses a sequence of state-action pairs of an expert for approximating a solution towards learning the policy (Pomerleau, 1989). However, BC is known to poorly generalize from the given problem due to compounding errors and covariant shift (Ross and Bagnell, 2010; Ross et al., 2011). On the other hand, AL tends to reconstruct the reward function (Abbeel and Ng, 2004; Syed et al., 2008; Ho et al., 2016), but at high computational costs due to embedded reinforcement learning in the training loop.

Generative Adversarial Imitation Learning. One prominent work on AL is the Generative Adversarial Imitation Learning (GAIL) approach (Ho and Ermon, 2016), which objective is to learn the policy for a given complex task without estimating the reward function directly. GAIL is based on the originally introduced generative adversarial network (GAN) (Goodfellow et al., 2014) which consists of two networks, a generator as the end-policy estimator π and a discriminator that has to differentiate between the given real and synthesized inputs generated from π_E and π, respectively. Thus, GAIL is intimately connected to GAN, however with a newly introduced objective. Mathematically, the objective of a GAIL is defined as follows:

$$\min_{\pi} \max_{D \in (0,1)^{S \times A}} \mathbb{E}_\pi[\log D(s,a)] + \mathbb{E}_{\pi E}[\log(1 - D(s,a))] - \lambda H(\pi) \tag{1}$$

where π denotes the agent's policy, π_E the policy of the expert, D the discriminative classifier, which tries to distinguish state-action pairs generated by π and π_E respectively. The $H(\pi) \triangleq \mathbb{E}[-\log \pi(a|s)]$ denotes the γ - discount casual entropy of the policy π_θ as defined by Bloem and Bambos (2014). The objective of a GAIL is to learn the optimal reward policy based on its discriminator D reasoning during the comparison. Thus, the optimal policy π is achieved, once D reaches maximum uncertainty state. The GAIL framework addresses the problem of learning a policy from example expert behavior without interaction with the expert or access to reinforcement signal. It is a model-free approach of directly extracting a policy from data, similar to inverse reinforcement learning. For this purpose, it leverages a hybrid optimization approach by alternating between gradient steps to maximize (1), achieved by using Monte-Carlo estimation of policy gradients, and trust region policy optimization (TRPO) (Schulman et al., 2015) to minimize (1) with respect to the agent policy π.

InfoGAIL. The InfoGAIL system (Li et al., 2017; Hausman et al., 2017) is an extension to GAIL, that motivates our work and serves as baseline for comparative experimental performance evaluation. In InfoGAIL, the agent's policy estimation is derived from the mixture of expert trajectories, where a direct relation to the latent variable has been added as proposed by Chen et al. (2016). Thus, the extended policy objective given latent variable c resulted in $\pi(a|s,c)$, which is an approximation to the π_E, c denotes a tuple of the form $[x, 1-x]$, where $x \in [0, 1]$. In order to force the network to rely on the introduced latent variable, an information-theoretic regularization has been utilized, which states that there should exist a high mutual information between the latent variable and state-action pairs in generated trajectories. Thus, the model objective in (1) extended with a variational lower bound $L_1(\pi,Q)$ of the mutual information $I(c;\tau)$, where τ denotes trajectory, is as follows:

$$\min_{\pi,Q} \max_{D} \mathbb{E}_\pi[\log D(s,a)] + \mathbb{E}_{\pi E}[\log(1 - D(s,a))] - \lambda_1 L_1(\pi,Q) - \lambda H(\pi) \tag{2}$$

where λ_1 is the hyper-parameter for the information maximization regularization term. Hence, the variational lower bound, $L_1(\pi,Q)$, of the mutual information is given as:

$$L_1(\pi,Q) = \mathbb{E}_{c \sim p(c), a \sim \pi(\cdot|s,c)} [\log Q(c|\tau)] + H(c) \leq I(c,\tau) \tag{3}$$

where $Q(c|\tau)$ is an approximation of the true posterior $P(c|\tau)$. After introducing an augment-reward term

(Englert and Toussaint, 2018) to incorporate prior knowledge and the Wasserstein GAN Arjovsky et al. (2017) to overcome the vanishing gradient problem, the final objective is denoted as:

$$\min_{\theta,\psi}\max_{\omega}\mathbb{E}_{\pi_\theta}[D_\omega(s,a)]+\mathbb{E}_{\pi E}[D_\omega(s,a)]- \lambda_0\eta(\pi_\theta)-\lambda_1 L_1(\pi,Q)-\lambda H(\pi) \quad (4)$$

where $\eta(\pi_\theta)=\mathbb{E}_{s\sim\pi_\theta}[s_r]$ reflects tendency towards learning the desired behavior.

3 The InfoSalGAIL Approach

The InfoSalGAIL system enables the imitation of human-like walking behavior of pedestrians by means of navigating avatars in realistic traffic scenarios generated with virtual driving simulators. Simulated behavior of a human pedestrian is considered safe or unsafe and is dependent on the current position of the expert in relation to the defined traffic scene zones (cf. Figure 3). InfoSalGAIL solves the pedestrian navigation problem defined in Section 2.2 with the following modifications with respect to visual attention and rewards.

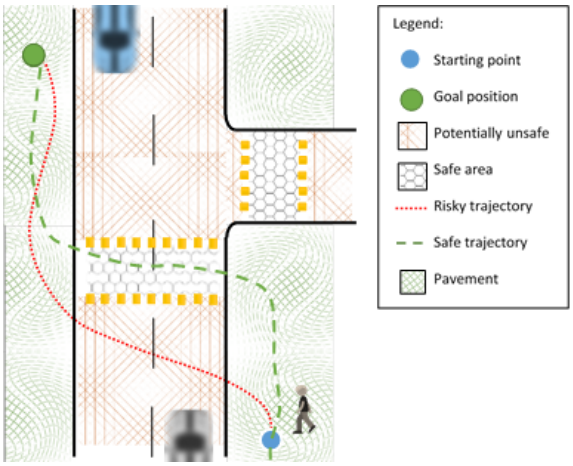

Figure 3: Example of the environment partitioned into scene zones. The red dotted trajectory shows a sample of risky navigation, where the green dashed one, highlights possible safe trajectory since zebra crossing taken to cross the street in pursuance of reaching the goal.

3.1 Visual Attention and Rewards

Visual attention and policy. The finite state space S of the considered MDP is extended with pairs of *visual* and *saliency* information (vis, sal). Thus, the expert policy is given in the form of visual demonstrations supported by the saliency heat maps described in Section 2.1 each of which stands for a fixation sequence. Thus, the expectation with respect to the policy π denotes an expectation with respect to the generated trajectories: $\mathbb{E}_\pi[f(s_{vis,sal},a)] \triangleq \mathbb{E}[\sum_{t=0}^{\infty}\gamma^t f(s_{t_{vis,sal}},a_t)]$, such that $s_0\sim\rho_0$, $a_t\sim\pi(a_t|s_{t\,vis,sal})$, $s_{t+1\,vis,sal}\sim P(s_{t+1\,vis,sal}|a_t,s_{t\,vis,sal})$.

Reward augmentation. As proposed by Englert and Toussaint (2018), a reward augmentation is required in order to gain prior knowledge about the environment as well as to compensate for a complex state space, e.g. urban scene, in which pedestrian avatar can literally walk (steer) to any direction. Thus, it is essential to introduce scene semantics as part of the surrogate reward function, which is defined as follows $r=(coll_{sta}, coll_{car}, nM_{car}, coll_{ped}, nM_{ped}, loc, dist_{goal}, vis_{goal})$, where

- $coll_{sta}\in]-1,1[$ denotes collision occurrence to the static objects of the scene, e.g. buildings, traffic signs, traffic lights, other street furniture, etc;
- $coll_{car}\in]-1,1[$ denotes collision occurrence concerning the vehicles;
- $nM_{car}\in]-0.5,0.5[$ stands for the occurrence of a near-miss with a reference to the vehicles;
- $coll_{ped}\in]-0.1,0.1[$ represents collision occurrence to any non-player character (NPC) pedestrian avatar;
- $nM_{ped}\in]-0.05,0.05[$ denotes the occurrence of a near-miss with reference to the NPC;
- $loc\in\{-1;0.5;1\}$ denotes the location of the avatar with respect to the defined simulation scene zones (cf. Figure 3);
- $dist_{goal}\in[0,1[$ denotes avatar's distance to the goal;
- $vis_{goal}\in]-1,1[$ stands for the in range of vision goal, e.g. in the case of an obstacle between the avatar and the goal, like a column in the FoV, the end-reward approaches -1.

The proposed surrogate reward adopts the label smoothing technique to discourage the discriminator from producing overconfident classification and ensuring that a much broader set of features considered during classification task, e.g. in the case of a near-miss with a vehicle, instead of using a fixed negative reward $[-0.5]$, a value in the range of $[-0.45;-0.55]$ is sampled. We adapt the near-miss/hit concept for the car (Pusse and Klusch, 2019) as illustrated on Figure 4 with the shifted focus towards pedestrian. The shape

of the near-miss area is dynamically scaled to cover fast walking activity, e.g. stretched in walking direction in accordance to the avatar's speed and is equal to the hit area in the case of idle. Due to the applied augmented reward function (similar to the baseline), InfoSalGAIL can be seen as a hybrid between reinforcement and imitation learning. Thus, the reinforcement signal for the purpose of policy optimization is a compound of explicitly defined surrogate reward and implicitly derived reward from the Discriminator.

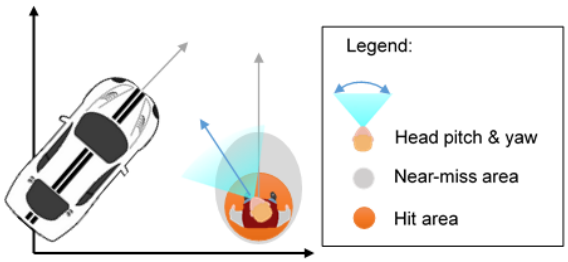

Figure 4: Sample of near-miss and hit areas. The grey vector indicates the move direction of both the vehicle and pedestrian. The blue vector indicates look direction of the avatar's head, where pitch and yaw actions are supported.

3.2 Saliency Generator

Inspired by the approach for visual saliency prediction with GAN (SalGAN) by Pan et al. (2017), InfoSalGAIL adopts the generator part of SalGAN that follows its convolutional encode-decode network architecture. This network is identical to the VGG-16 proposed by Simonyan and Zisserman (2014) network, where the last five layers are removed to update the training objective from soft-classification task to match reconstruction of the input image objective. The initial weights of the network initiated with pre-trained weights of VGG-16 given ImageNet classification dataset, to achieve feature extraction. In addition to the generator architecture, we added a Dropout layer after the second cascade of convolutional layers of decoder to avoid, experimentally confirmed, model over-fitting. During the training, the batch size is set to 96, where Adagrad (Duchi et al., 2011) optimizer was utilized with the loss rate of 0.0001.

The loss is computed on a per-pixel basis, based on the binary cross-entropy function and is denoted as:

$$\Delta_E = -\frac{1}{N}\sum_{j=1}^{N}(M_j \log(\hat{M}_j) + (1-M_j)\log(1-\hat{M}_j)) \quad (5)$$

where $\hat{M}$ stands for the predicted saliency value (map) and M denotes the ground truth respectively, $N = W_{img} \times H_{img}$ is the resolution of the input image. The sequence of image-saliency triples on Figure 5 demonstrates Case-2 critical scenario, based on GIDAS analysis, given the crossing the street objective. As illustrated, the left column refers to the avatar's FoV, where the middle column denotes the ground-truth saliency map, additionaly extended with a short-term visual attention memory build upon the corresponding saliency frames from $t-1$ and $t-2$ (circular areas of smaller diameter). The rightmost column refers to the generated saliency maps.

3.3 System Architecture

The InfoSalGAIL system architecture consists of four different networks that are trained separately:

- the saliency generator network $\Delta_E(s)$ as in Section 3.2 with the objective to reconstruct expert's most probable visual attention map;
- the extended policy generator network $\pi_\theta(a|s_{vis,sal},c)$ (cf. Figure 6) which corresponds to the end-learning policy we would like to approximate;
- the extended discriminator network $D_\omega(s_{vis,sal},a)$ (cf. Figure 10), with the objective to differentiate between the synthetic and true (expert) inputs;
- the extended posterior estimator network $Q_\psi(c|s_{vis,sal},a)$ (cf. Figure 11), with the objective to reconstruct the latent variable, in particular, safe and unsafe navigation styles, given visual demonstrations.

In contrast to the InfoGAIL work, where RMSProp gradient descent algorithm is applied, D_ω is updated utilizing Stochastic Gradient Descent (SGD) optimizer with Nesterov momentum and $lossrate = 1e^{-6}$, resulting in slower policy convergence, allowing for better problem generalization given substantial state-space and variations. The update of Q_ψ and π_θ performed by alternating between Adam optimizer (Kingma and Ba, 2014) and TRPO as proposed by Schulman et al. (2015). In accordance with Ho and Ermon (2016), to accelerate network convergence, the weights of π_θ initialized through the behavior cloning (BC) pre-trained network, enhanced by usage of visual attention information. As in original InfoGAIL, the discriminator network D_ω and the posterior approximation network Q_ψ are threaded as different networks, due to the applied weight clipping and momentum-free optimization methods in the process of training D_ω (to avoid interference with Q_ψ).
After applying all the extensions, the final training ob-

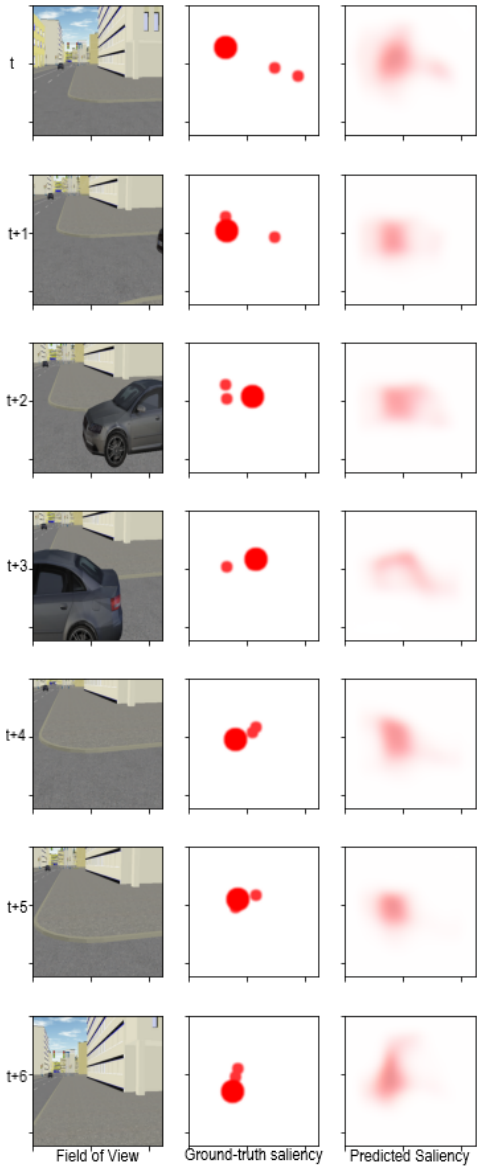

Figure 5: Sample of a saliency generator input-output. Performed on a continuous sequence of 7 images, given the crossing of the street objective, with 10 Hz interval (top to bottom order). Left column: FoV of the avatar; Middle column: blurred saliency ground-truth map; Right column: predicted saliency map. The achieved performance is equal to 0.06261 ($> 6\%$), given an input dataset of 1K images, with the training/testing ratio of 0.7.

jective is given as:

$$\min_{\theta,\psi,\Delta} \max_{\omega} \mathbb{E}_{\pi\theta}[D_\omega(s_{vis,sal}, a)] + \mathbb{E}_{\pi E}[D_\omega(s_{vis,sal}, a)] - \lambda_0 \eta(\pi_\theta) - \lambda_1 L_1(\pi, Q) - \Delta_E(s) - \lambda H(\pi) \quad (6)$$

Training and Inference. As shown on Figure 6, the policy estimator π_θ accepts two elements as an input: the input image of the size $224 \times 224 \times 3$ and a set of XML files passed over in unpacked vector format as auxiliary information for the frames at time $t-1$ and $t-2$, denoting short-term memory. Concurrently, the input image pass on to saliency generator Δ_E to derive the most probable visual attention map for the current frame. The input image and obtained saliency map are then feed through the ResNet50 (up to activation layer 40) network followed by a cascade of convolution layers to extract valuable feature maps. Eventually, the feature maps of both signals are merged through the average operation resulted in drawing more attention to the features with higher prior visual attention information (saliency map). As depicted on Figure 6, afterward the flatten vector is concatenated to the prior knowledge (memory) followed by a set of fully connected layers merged (sum) with the latent variable, which denotes the behavior objective style, namely safe or unsafe on-street navigation.

The discriminator D_ω and posterior estimator Q_ψ networks both accept the same input, which is the input image and saliency map pair, auxiliary information, and the action vector derived by the π_θ on the previous step. However, the objective of the two is different, the Discriminator, as shown on Figure 10, uses this information to estimate the origin of the data by comparing synthetically generated input with the true (expert) input. The Posterior, on the other hand, aims for the latent variable (behavior class) prediction (cf. Figure 11 in Appendix) given a pair of input images and corresponding saliency maps.

During the inference, the saliency generator Δ_E and policy estimator π_θ are the only involved networks.

4 Experiments

Dataset creation. A set of relevant critical scenarios is a prior must-have condition before proceeding with the numerical evaluation. Thus, we created, first of its kind, pedestrian-centric dataset of the relevant critical scenarios, named OpenDS-GIDAS Motivated dataset (OpenDS-GiM)[3] based on the German In-Depth Accident Study analysis, which consists of pairs of visual demonstrations and corresponding saliency maps. In particular, the generated dataset consists of 9 classes of the most common critical situations visualized on Figure 7, recorded using an open source simulation software OpenDS[1] (given a beforehand created virtual twin of "name omitted"

[3] https://cloud.dfki.de/owncloud/index.php/s/XarwdHgDYYmma7H

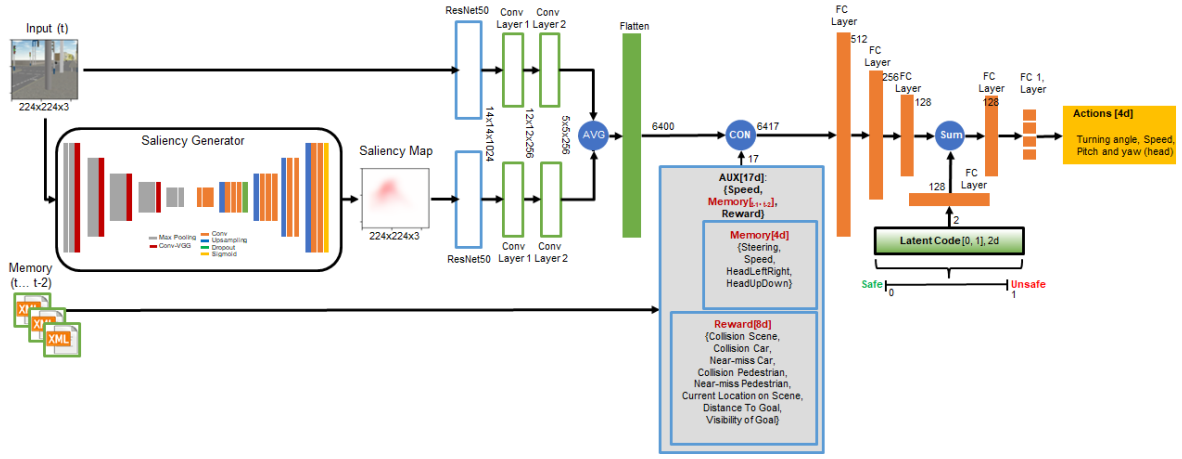

Figure 6: Visual Attention-Empowered Policy Generator architecture. The input image $(224 \times 224 \times 3)$ is passed to both ResNet50 (for feature extraction) and Saliency Generator (for saliency map generation). Following the cascade of convolutional layers, both inputs are merged by using the AVG operation, resulted in highlighting more valuable feature maps. The memory (XML) files are used as auxiliary information during the training and inference. Upon merging flatten feature maps with auxiliary information and latent variable, an output vector is generated. The output of the network (4d vector) denotes the turning angle (converted to a walking vector in OpenDS) and speed of the avatar as well as the pitch and yaw dimensions of the movement of the head. The latent code, in the case of inference, is used to interpolate between selected behavior objectives safe and unsafe.

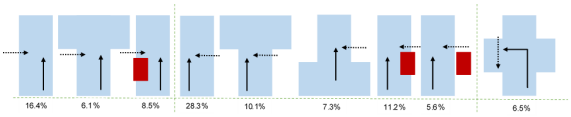

Figure 7: Layout of GIDAS on-street critical scenarios. The nine classes of accident scenarios between the car and the pedestrian, clustered in accordance to the street layout with the corresponding scenario occurrence ratio in percentage. The dotted vector denotes pedestrian moving direction, where the solid vector states for the vehicle moving direction. The red rectangle is an obstacle, e.g. vehicle.

city, Germany) supported by an eye-tracker sensor for accessing the true expert fixation sequences. Prior to the start of the recording session, an eye tracker has been calibrated for every expert (6 in total), where the overall precision of the chosen eye-tracker falls under 0.7 degrees, and recall under 0.25 degrees. In total, the released dataset[3] contains 140K of pairs of FoV images and corresponding saliency maps.

Trajectory Generation. Due to the existence of potentially endless set of possible avatar trajectories in the simulation environment, the variance reduction technique, in particular the common random numbers (CRN) method has been applied. In contrast to the vine path generation approach, initially proposed by Schulman et al. (2015) in Section 5.2, the initial starting point and orientation of the avatar is subject to a normal distribution $N(0,1)$. Thus, the trajectory generation workflow starts with sampling $s_n \sim \rho_0$ initial states denoted by the tuple $(Position, Rotation)$ and follows with simulating the policy π_θ, bounded by CRN (cf. Figure 8). The generated trajectories additionally bounded by a certain length, through a fixed time horizon, successfully achieved target objectives, and/or false navigation, e.g. throughout safe trajectory generation leaving of a safe zone leads to a negative reward and simulation reset. Hence, the generated actions, influenced by the CRN looks as follows: $a_n = \pi(\cdot|S_n, \sigma)$.

4.1 Results

InfoSalGAIL framework is relying on the visual attention information from the current frame t, which was additionally extended with the reduced visual attention map from frames $t-1$ and $t-2$ to compensate for the gaps between the two sequential visual inputs (frames are captured and passed over to the neural network framework at 10Hz rate resulted in only two full frames per second). The auxiliary information input for the policy generator at time t consists of a 17-dimensional vector with following elements: 1) **speed** $= 1d$ at time t; 2) **actions** $= 4d$ at time $t-1$ and $t-2$, compound of (turning angle, speed, pitch, and yaw); 3) **reward** $= 8d$ (cf. Section 3.1) at time t, where d denotes the dimension of elements. In the scope of this work, we considered two classes of on-street behavior: **safe**, where as demonstrated by the expert, the objective is to cross the street and reach the goal area by utilizing (pedestrian) safe zones in the scene, e.g. pavement/sidewalk, zebra crossings;

Table 1: Comparison evaluation of InfoGAIL and InfoSalGAIL frameworks given the street crossing task (Cases 1-3 based on GIDAS analysis, conditioned by approaching vehicle from the right). InfoSalGAIL outperformed the baseline in all categories, except for *collision to street furniture* one. Moreover, in the case of the **safe** simulation, in 50% of simulated scenarios, avatar followed zebra section of given street layout to cross the street (resulted in a fraction of 0.269 out of all trajectory points belonging to drivable scene zones, e.g. parking lots, pedestrian crossings, and drivable lanes). During the simulation, an average value across all simulated trajectories, has been chosen as evaluation metrics, where a single trajectory consists of a set of points, denoted by X and Y coordinates.

Approach	**InfoGAIL (baseline)**		**InfoSALGAIL (ours)**	
Targeted behavior style of navigation:	Safe	Unsafe	Safe	Unsafe
Target area (goal) reached following the chosen objective behavior style, e.g. safe, unsafe? [higher better]	0.0	0.0	**0.5**	**0.86**
Objective (street crossing) reached? [higher better]	0.0	0.0	**1.0**	**1.0**
Collision to the street furniture (buildings) occurred? [lower better]	**0.0**	**0.0**	0.056	0.07
Collision to the vehicles occurred? [lower better]	0.05	0.035	**0.045**	**0.03**
Near-miss to the vehicles occurred? [lower better]	0.054	0.045	**0.01**	**0.04**
Avatar remained within the pavement area? (fraction out of all trajectory points) [higher better]	0.345	0.366	**0.799**	**0.742**
Avatar remained within drivable area? (fraction out of all trajectory points) [lower better]	0.655	0.633	**0.201**	**0.258**
Avatar navigated through the pedestrian crossing? (fraction out of drivable area trajectory points) [higher better; not relevant for unsafe]	0.112	–	**0.269**	–

unsafe uses same objective, however, with no limitations with reference to scene zones, e.g. street lanes, aside parking lots are allowed for navigation. Nevertheless, the time constraint is not considered to be a decisive factor due to the significant variability of input data, e.g. the same expert can demonstrate different time to reach the goal performance as the result of the scene dynamics or other latent factors. Thus, it is not part of the evaluation schema.

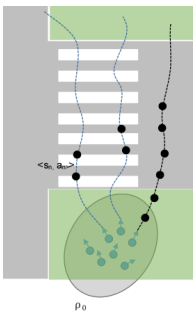

Figure 8: Trajectory generation approach, where ρ_0 corresponds to the initial distribution of starting points in S, where $s = (Pos_{x,y,z}, Rot_{x,y,z})$. The adopted CRN technique aims to reduce the variance of the data, where the blue dashed lines denotes samples of a *safe* trajectory, and black dashed one correspond to an *unsafe* one, respectively. The tuple $< s_n, a_n >$ represents a sample of a state-action pair along the generated trajectory.

Since the considered problem state-space is multi-dimensional, it is a challenge on its own to mathematically define relevant evaluation metrics, e.g. one-to-one trajectory comparison (expert vs avatar) cannot be considered. To overcome this limitation, we propose the following objective-oriented evaluation as summarized in Table 1 and visually demonstrated in Figure 9, where the trajectories of safe and unsafe navigation styles for both InfoGAIL and InfoSalGAIL frameworks as well as the expert demonstrations (ground truth) are plotted. As shown in the plots, the substantial increase in state-space led to a poor objective generalization by the baseline: the InfoGAIL model was not able to cope with the given task. Moreover, the desired target task of pedestrian agent to cross the street was never reached, despite different latent variable inputs policy generator $\pi_{\theta InfoGAIL}$ produced rather similar end-trajectories. The very same is shown in Figure 12 (middle row) in the Appendix, where we applied a dimension reduction technique, namely Principal Component Analysis (PCA) (Jolliffe, 2003), to visualize the policy generator output vector (in less dimensions) That aimed at controlling the avatar within simulated environment by means of principal components and corresponding variations. The PCA plots (middle row, Figure 12) confirms high similarity with actions produced by the policy generator $\pi_{infoGAIL}$ for the chosen walking styles, e.g. safe, unsafe. This can serve as another proof of poor model generalization given human-like on-street trajectory generation task. In contrast, trained InfoSalGAIL models reached the target objectives, namely street crossing and navigation to the *Goal* area, in $92,3\%$ of simulated scenarios. Moreover, in $\sim 50\%$ of **safe simulated** scenarios, despite unbalanced input data distribution (on average, a single complete trajectory contains $\sim 15\%$ of crosswalk visual demonstrations) the trained model even managed to mimic an exact expert style of navigation, e.g. street crossing through crosswalk. In the remaining simulated safe cases, the shortest path to cross the street while taking into the consideration approaching vehicles (head turns towards vehicles) has been generated by $\pi_{\theta InfoSalGAIL}$ (which might also be considered as a subclass of a safe navigation). The PCA plots (given the policy generator output vector) on Figure 12 (upper vs bottom rows)

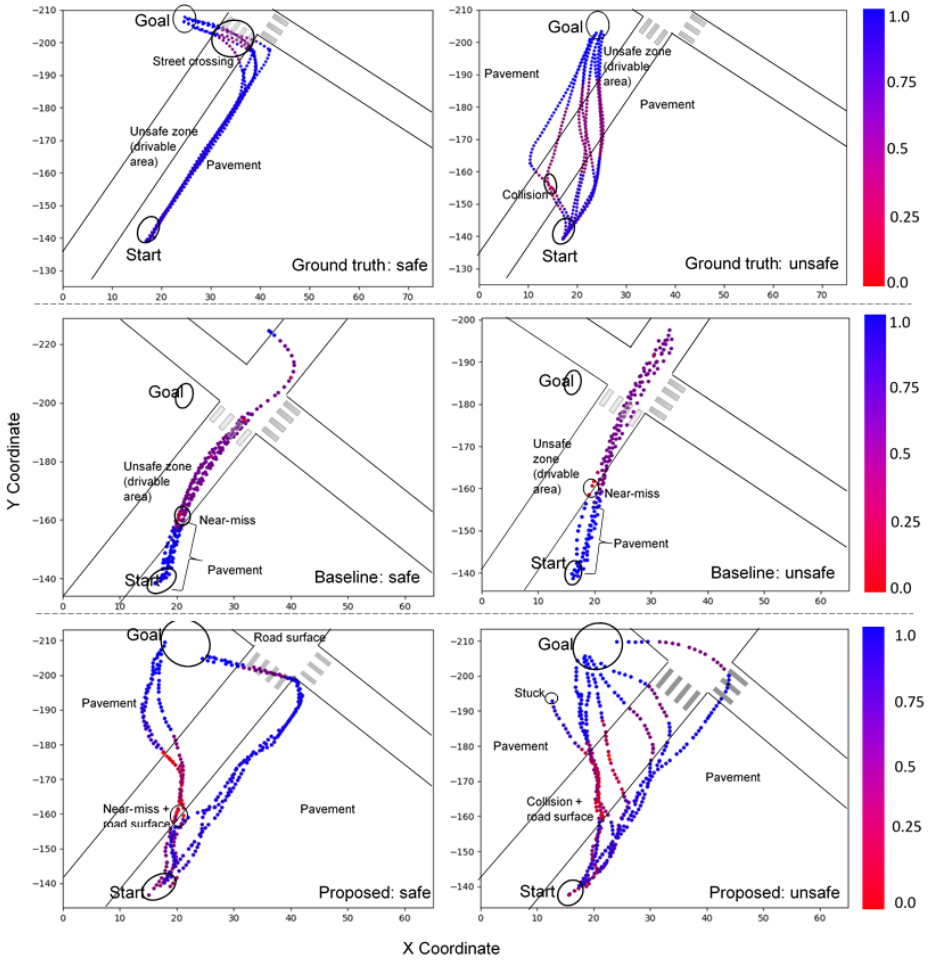

Figure 9: Plots of expert trajectories gathered during visual demonstrations (top row) together with policy generated trajectories produced by InfoGAIL (baseline, middle row) and InfoSalGAIL (proposed, bottom row) frameworks. The scene elements, e.g. vehicles, buildings, parking lots as well as simulation specific details like pitch and yaw (head turns) were excluded from the plots for visual transparency. The color bar on the right denotes the normalized surrogate reward. As seen from the plots, the baseline fails to generalize given problem objective resulting in generating rather similar trajectories for safe and unsafe navigation styles. Moreover, it demonstrates poor performance in reaching the goal area and street crossing task. In contrast, the trajectories generated by the proposed model are comparable to the expert trajectories.

in Appendix confirms the high similarity between the principal components of ground truth and generated by InfoSalGAIL framework data, where the numerical representation of principal components and corresponding variations are summarized in Table 2 in Appendix. Nevertheless, the collision with a vehicle still occurred. Such an anomaly can be caused by the input data distribution used in the process of training of saliency generator, where a dataset including (evenly distributed) accident cases is required for more accurate visual attention map estimation. In the case of **unsafe simulation**, the policy generator $\pi_{\theta InfoSalGAIL}$ reached the street crossing target in 100% of simulated cases. However, only in $> 86\%$ of simulated scenarios, the goal area has been reached, where in the process of simulation both the near-miss as well as the collision with the vehicles took place (indicated by negative surrogate reward explicitly obtained from simulated environment). To support plotted trajectories (cf. Figure 9) given GIDAS scenarios Cases 1-3 (crossing the street with an approaching vehicle from the right), in a more intuitive manner, a video[4], for the purpose of performance demonstration of InfoGAIL and InfoSalGAIL, has been recorded. Additionally, avatar's FoV augmented with the visual attention map was compiled into an animated GIF[4].

The conducted experiments reveal that InfoSal-

[4]https://www.dropbox.com/sh/smm2vxbuwwlctez/AAD2AmcZ9kZjMEAeiHV3WMHaa?dl=0

GAIL outperformed InfoGAIL by utilizing additional input information, namely expert's visual attention map. Moreover, since the visual attention map is unique for every expert, it might characterize each individual and serve as additional finger-print like feature.

The entire framework training took place by utilizing a high compute server with NVIDIA Tesla V100 GPU (32GB), where the CPU-memory consumption was roughly 90GB, due to the in memory loaded dataset of images. InfoSalGAIL framework is based on TensorFlow (version 1.15) and Keras 2.0 library, where the connection to OpenDS has been realized though a transmission control protocol (TCP) to guarantee no data loss.

5 CONCLUSIONS

In this paper, we presented a novel approach, named InfoSalGAIL, for visual attention-empowered imitation learning of pedestrian behavior in critical traffic scenarios that can handle substantial state-space and variations, e.g. on-street urban scenarios, to mimic complex human-like behavior of experts in a virtual environment. Moreover, we synthesised two classes of navigation (cf. in Section 3) which renders InfoSalGAIL quite suitable for the challenge of critical traffic scenario generation. Our experimets revealed that InfoSalGAIL can significantly outperform the selected baseline InfoGAIL for the given objective due to the utilization of a saliency map and its direct impact on the policy generator in deriving the output vector (control actions). To support this research activity, the functionality of the OpenDS simulation software has been extended to allow for a pedestrian-centric control, resulting in a creation of a new dataset, which consists of more than 140K pairs of images and corresponding saliency maps generated from a virtual clone of Saarbruecken city (Germany).

Future research is concerned with an extension of the saliency generator network by incorporating latent variables to further differentiate between the pedestrian imitating avatars such in terms of age, average speed, short term interests. In this regard, the created benchmark will be extended with a new set of realistic scenarios based on JAAD[5] dataset to capture ground truth data. In general, we hope that InfoSalGAIL attracts more attention to the topic of human-like behavior simulation in the scope of generating critical traffic scenarios for virtual tests and validation of collision-free navigation methods of self-driving cars.

[5]http://data.nvision2.eecs.yorku.ca/JAAD_dataset

ACKNOWLEDGEMENTS

This research was funded by the German Federal Ministry for Education and Research (BMBF) in the project REACT under grant 01IW17003.

REFERENCES

Abbeel, P. and Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.

Bloem, M. and Bambos, N. (2014). Infinite time horizon maximum causal entropy inverse reinforcement learning. In *53rd IEEE Conference on Decision and Control*, pages 4911–4916. IEEE.

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., and Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in neural information processing systems*, pages 2172–2180.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences*, 95(3):831–838.

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).

Englert, P. and Toussaint, M. (2018). Inverse kkt–learning cost functions of manipulation tasks from demonstrations. In *Robotics Research*, pages 57–72. Springer.

Ermon, S., Xue, Y., Toth, R., Dilkina, B., Bernstein, R., Damoulas, T., Clark, P., DeGloria, S., Mude, A., Barrett, C., et al. (2015). Learning large-scale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in east africa. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Finn, C., Levine, S., and Abbeel, P. (2016). Guided cost learning: Deep inverse optimal control via policy optimization. In *International conference on machine learning*, pages 49–58.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Hausman, K., Chebotar, Y., Schaal, S., Sukhatme, G., and Lim, J. J. (2017). Multi-modal imitation learning from

unstructured demonstrations using generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 1235–1245.

Ho, J. and Ermon, S. (2016). Generative adversarial imitation learning. In *Advances in neural information processing systems*, pages 4565–4573.

Ho, J., Gupta, J., and Ermon, S. (2016). Model-free imitation learning with policy optimization. In *International Conference on Machine Learning*, pages 2760–2769.

Jolliffe, I. (2003). Principal component analysis. *Technometrics*, 45(3):276.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, Y., Song, J., and Ermon, S. (2017). Infogail: Interpretable imitation learning from visual demonstrations. In *Advances in Neural Information Processing Systems*, pages 3812–3822.

Pan, J., Ferrer, C. C., McGuinness, K., O'Connor, N. E., Torres, J., Sayrol, E., and Giro-i Nieto, X. (2017). Salgan: Visual saliency prediction with generative adversarial networks. *arXiv preprint arXiv:1701.01081*.

Pomerleau, D. A. (1989). Alvinn: An autonomous land vehicle in a neural network. In *Advances in neural information processing systems*, pages 305–313.

Pusse, F. and Klusch, M. (2019). Hybrid online pomdp planning and deep reinforcement learning for safer self-driving cars. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 1013–1020. IEEE.

Ross, S. and Bagnell, D. (2010). Efficient reductions for imitation learning. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 661–668.

Ross, S., Gordon, G., and Bagnell, D. (2011). A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015). Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Stadie, B. C., Abbeel, P., and Sutskever, I. (2017). Third-person imitation learning. *arXiv preprint arXiv:1703.01703*.

Sully, J. (1891). W. james, the principles of psychology.

Syed, U., Bowling, M., and Schapire, R. E. (2008). Apprenticeship learning using linear programming. In *Proceedings of the 25th international conference on Machine learning*, pages 1032–1039.

Ziebart, B. D., Maas, A. L., Bagnell, J. A., and Dey, A. K. (2008). Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA.

APPENDIX

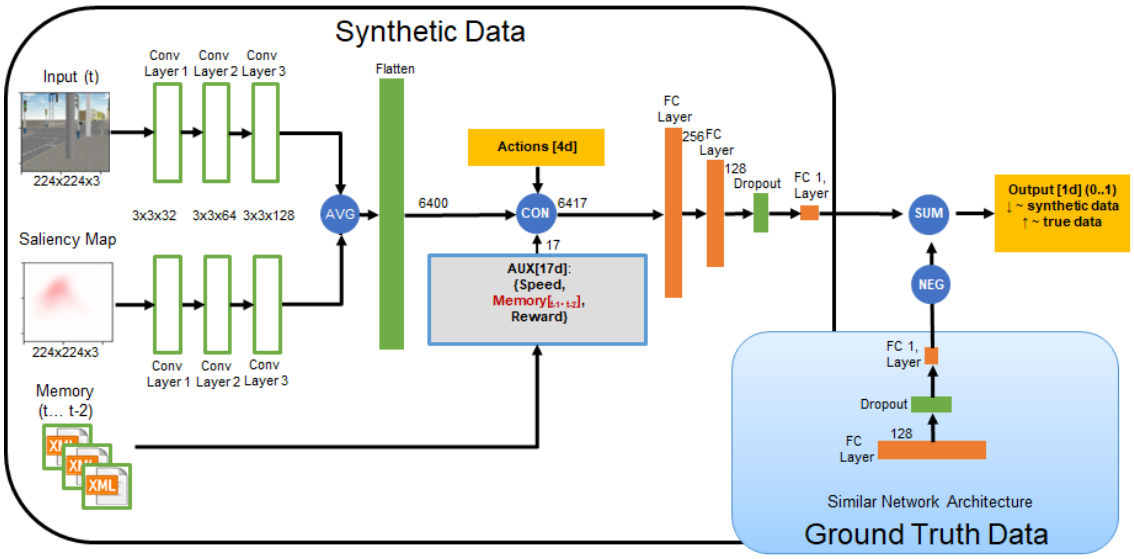

Figure 10: Visual Attention-Empowered Discriminator Network architecture. The input image together with the corresponding saliency map of the size $224 \times 224 \times 3$ are passed through a set of convolutional layers in order to derive feature maps for later merging (AVG). Thus, features empowered with the saliency map become more valuable. The memory, represented in the form of XML files are used as auxiliary information in the process of training, where $4d$ vector of actions generated by the Policy generator is given as additional input. The ground truth data, obtained during the dataset creation phase (recordings of expert visual demonstrations and corresponding saliency maps), passed through the identical network to identify the origin of the input (synthetic or ground truth). The goal is to reach the maximum uncertainty state as an output of Discriminator network.

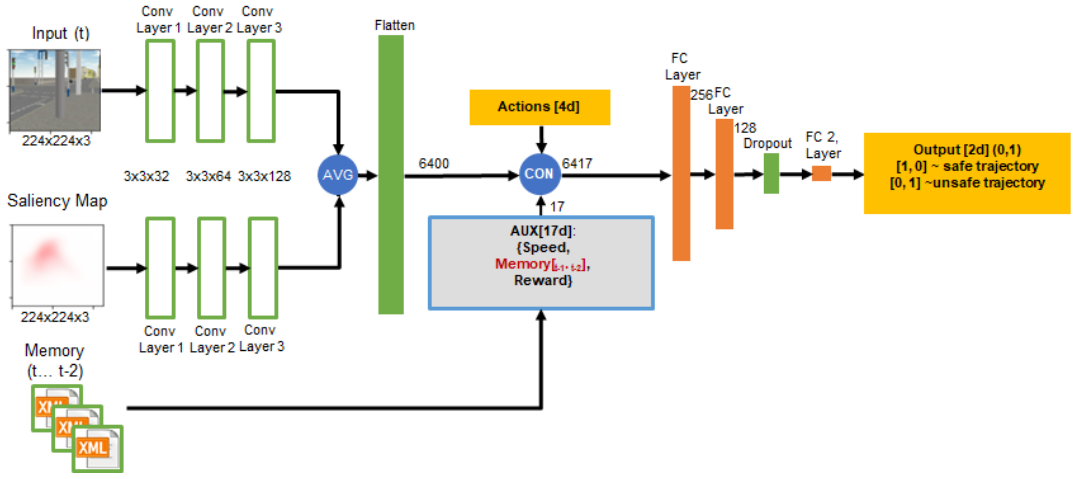

Figure 11: Visual Attention-Empowered Posterior Estimator Network architecture. The input image of size $224 \times 224 \times 3$ together with the saliency map are passed through a set of convolutional layers in order to derive feature maps for later merging (AVG). Thus, features empowered with the saliency map become more valuable. The memory, represented in the form of XML files are used as auxiliary information in the process of the training, where $4d$ vector of actions generated by the Policy generator is given as additional input. The objective of this network is to derive latent variable (navigation style, e.g. safe or unsafe) without explicitly proving an optimal reward function since it is a challenge on it's own.

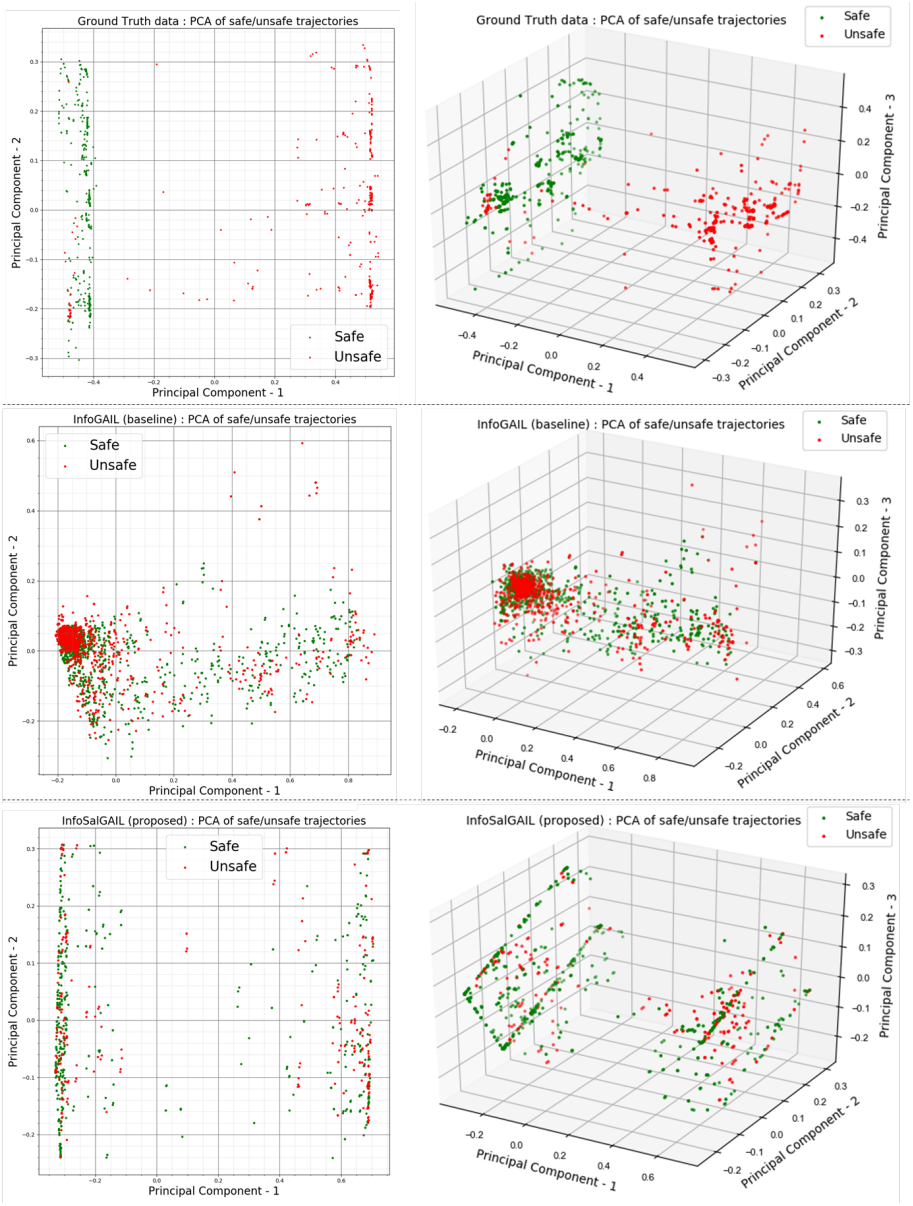

Figure 12: The numerical evaluation of the output vector using dimension reduction approach of PCA. Top row: expert ground truth data; middle row: InfoGAIL (baseline) generated data; bottom row: InfoSalGAIL (proposed) generated data. The red dots characterise unsafe navigation style, where green dots represent a safe one. As seen from InfoGAIL plots, the model fails to generalize problem objective. Thus, resulting in generating very similar trajectories for both types of navigation. In contract, the PCA visualization of the output vector of InfoSalGAIL policy generator looks comparable to the ground truth data, which is confirmed from the plotted trajectories in Figure 9.

Table 2: Variation per principal component given output vector, e.g. turning angle, speed, pitch, and yaw.

Data / Principal Component	**1**	**2**	**3**
Ground Truth (expert)	0.81556843	0.09826111	0.06400615
InfoGAIL (baseline)	0.87912261	0.08173591	0.02932816
InfoSalGAIL (proposed)	0.8115204	0.10881166	0.06633512