

---

# Controlling Steering with Energy-Based Models

---

**Anonymous Author(s)**

Affiliation  
Address  
email

## Abstract

1 So-called implicit behavioral cloning with energy-based models has shown promising  
2 results in robotic manipulation tasks. We tested if the method’s advantages  
3 carry on to controlling the steering of a real self-driving car with an end-to-end  
4 driving model. We performed an extensive comparison of the implicit behavioral  
5 cloning approach with explicit baseline approaches, all sharing the same neural  
6 network backbone architecture. Baseline explicit models were trained with regres-  
7 sion (MAE) loss, classification loss (softmax and cross-entropy on a discretization),  
8 or as mixture density networks (MDN). While models using the energy-based  
9 formulation performed comparably to baseline approaches in terms of safety driver  
10 interventions, they had a higher whiteness measure, indicating higher jerk. To  
11 alleviate this, we show two methods that can be used to improve the smoothness of  
12 steering. We confirmed that energy-based models handle multimodalities slightly  
13 better than simple regression, but this did not translate to significantly better driving  
14 ability. We argue that the steering-only road-following task has too few multi-  
15 modalities to benefit from energy-based models. This shows that applying implicit  
16 behavioral cloning to real-world tasks can be challenging, and further investigation  
17 is needed to bring out the theoretical advantages of energy-based models.

18 **1 Introduction**

19 Implicit behavioral cloning[14] with energy-based models [22] has shown a lot of promise in robotic  
20 manipulation tasks. The theoretical advantages of energy-based models include increased data  
21 efficiency and the ability to model discontinuities and multimodalities in the output action distribution  
22 [14]. Here, we set out to evaluate energy-based models for controlling the steering of a self-driving  
23 vehicle using end-to-end driving models [27, 5]. We work with steering-only models as the usefulness  
24 of multimodality-handling is evident in this output modality, as illustrated by the theoretical and  
25 practical failure cases of unimodal models on Figure 1. Adding longitudinal control would complicate  
26 the task and demand more training data while not necessarily adding more multimodal situations.  
27 To validate the theoretical advantages empirically, we compare a simple energy-based model (EBM)  
28 with several baseline approaches in a road-following task in the real world. The main experiments  
29 were also repeated in the VISTA[2] simulator. The explicit baseline models are based on the same  
30 neural network architecture as the EBM, with only the necessary modifications. They are trained  
31 with the MAE loss, classification loss (softmax with cross-entropy), or as mixture density networks  
32 [4] (MDN). During on-policy testing, the models controlled only the car’s steering; the location- and  
33 direction-relevant velocity was taken from a previously recorded expert trajectory. The evaluation



Figure 1: Left: If the experts have passed the tree from left and right with equal frequency in the training data, behavioral cloning with a unimodal policy and no high-level navigation commands would average the training trajectories and drive straight into the tree. Right: In practice, we have experienced such behavior at locations where side roads enter the main road - unimodal regression models tend to swerve slightly (the red trajectory) towards the side road. These swerves are minor because side roads are rarely taken in our training data, and keeping straight is the dominant behavior.

34 was performed on a WRC 2022 Rally Estonia track designed to be challenging for humans, which  
35 was not included in the training set.

36 According to the main evaluation metric, safety-driver intervention count, energy-based models  
37 performed comparably to the baseline methods but had noticeably jerkier steering. To alleviate this,  
38 we proposed two methods: temporal smoothing of predicted steering angles and spatially-aware soft  
39 targets for cross-entropy loss. Still, unimodal explicit behavioral cloning performed best in terms of  
40 safety driver interventions and jerk. We argue that the inductive bias enforced by unimodal losses  
41 makes them more data-efficient in simpler road situations, but more data is needed to model situations  
42 requiring true multimodalities.

43 The main contributions of the paper are as follows:

- 44 **1.** We show that controlling the steering of an autonomous vehicle with energy-based models in the  
45 real world performs comparably to the baseline explicit behavioral cloning approaches.
- 46 **2.** We propose two methods that effectively reduce the steering jerk of energy-based models: temporal  
47 smoothing of predicted steering angles and soft targets for the cross-entropy loss.
- 48 **3.** To our surprise, we find that energy-based models do not outperform any of the similar-architecture  
49 explicit behavioral cloning baseline approaches in the real-world road-following task and that repre-  
50 senting multimodalities does not translate into better driving.

## 51 **2 Background**

52 **End-to-end driving** End-to-end driving attempts to replace the classical modular self-driving  
53 pipeline with a single neural network model [27]. In the purest form, an end-to-end self-driving  
54 model takes in raw sensor data and yields actionable commands such as steering angle, throttle, and  
55 brake values. Such models are commonly trained with behavioral cloning [26, 5] to imitate human  
56 expert commands in the same situation.

57 One of the popular end-to-end driving models is NVIDIA PilotNet [5]. In our work, we use the  
58 PilotNet network architecture as the backbone of all models because it is relatively fast to train  
59 and sufficient for the road-following task we aimed for. We do not use conditioning on high-level  
60 commands [8] or more complex network architectures [9] to keep the setup simple - our goal is to  
61 compare similar-capacity implicit and explicit models, not to aim for the best performance. We use  
62 steering angle as the network output, rather than trajectory[6, 15] or costmap[29], to be in line with  
63 prior implicit learning work in robotics [14] where the network predicted raw actuator signal.

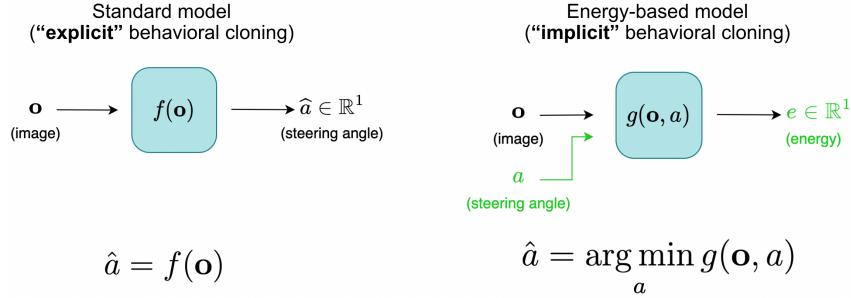


Figure 2: Controlling steering with implicit behavioral cloning. Left: explicit models output the predicted angle directly. Right: implicit energy-based models return an energy value per observation-action pair. The action yielding the lowest energy for the current observation is chosen via *argmin*.

64 A significant amount of end-to-end driving research is done in simulations[10], which through  
 65 repeatability and access to the state of the world, allow benchmarking with more complex measures  
 66 of driving quality [9, 1]. However, models created in simulations cannot be deployed in the real world  
 67 without adaptations [3]. Also, some problems related to steering delays and passenger comfort only  
 68 become apparent in the real world. Here, we have sufficient data for learning the relatively simple  
 69 road-following task, so we choose to work in the real world.

70 While the data-driven approach to autonomous driving is viewed as the most promising path to full  
 71 autonomy by some authors [18, 16], significant problems remain to be solved, most prominently in  
 72 generalization, in the explainability of decisions, and in providing safety guarantees [9, 24, 19].

73 **Energy-based models** An *energy function*, described by Lecun et al. [22], is any continuous  
 74 function that measures "goodness" between two sets of variables, where "good" pairs have a low  
 75 energy value. Following Florence et al. [14] who coined the term, we call behavioral cloning policies  
 76 "implicit" when they are composed of *argmin* and a continuous energy function  $E$ , such that:

$$\hat{a} = \arg \min_{a \in \mathcal{A}} E(\mathbf{o}, a), \quad (1)$$

77 where  $\mathbf{o}$  is an observation, e.g., a camera image, and  $a$  is an action, e.g., a steering wheel angle.  
 78 In the present work,  $E$  is implemented by a neural network with PilotNet architecture with minor  
 79 modifications (discussed below). The classical approach of a model directly computing the action  
 80 based on an observation is in this context called "explicit" behavioral cloning (Figure 2).

81 Implicit behavioral cloning with energy-based models promises the following three advantages  
 82 compared to classical explicit behavioral cloning: the ability to represent discontinuities sharply,  
 83 the ability to represent multimodal action distributions, and better generalization and improved data  
 84 efficiency. In this work, we mainly focus on the ability to model multimodalities.

85 **Evaluation of driving models** With behavioral cloning, models are optimized to make momentary  
 86 decisions on data originating from the distribution resulting from human driving. When deployed,  
 87 however, the solutions face a sequential decision-making task on data originating from a distribution  
 88 caused by their own driving. Off-policy metrics computed on held-out datasets of expert driving,  
 89 such as mean absolute error (MAE), measure only the predictive ability of the models. However, such  
 90 measures are insufficient for predicting success at the sequential decision-making task when deployed  
 91 [7]. Despite modest correlations with driving ability, we use these metrics for model selection, as is  
 92 often done in related works.

93 Among the on-policy metrics measured during model deployment, the number of safety driver  
 94 interventions, the mean distance between interventions (DBI), and the amount of time or distance  
 95 traveled autonomously are the most popular metrics [27]. In our work, we chose the number of  
 96 interventions as the main metric, as the distance traveled was fixed.

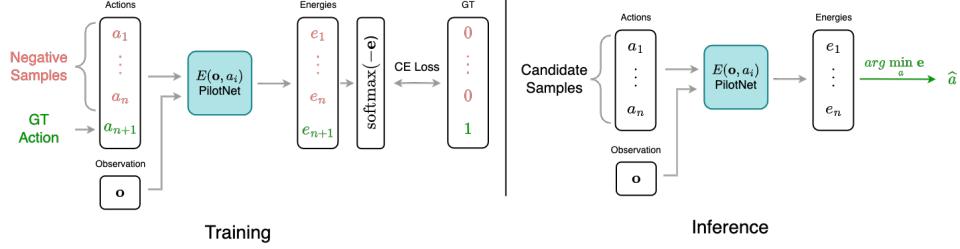


Figure 3: Energy-based model training and inference procedures. Left: feeding the observation and different action values (one action at a time) through the energy-computing network results in a vector of energy values, which is then optimized via CE loss to make the ground truth action have the lowest energy. Right: Observation is fed through the network with different candidate action values, producing an energy value per candidate; the lowest-energy action is chosen.

97 Beyond just completing routes safely, the comfort of passengers matters. The smoothness of driving  
 98 has been related directly to passenger comfort and perceived safety[17, 11]. We follow multiple  
 99 previous works [12, 13, 28] that quantify the smoothness of steering with *whiteness*, defined as:

$$W = \sqrt{\frac{1}{D} \sum_{i=1}^D \left( \frac{\delta P_i}{\delta t} \right)^2}, \quad (2)$$

100 where  $\delta P_i$  is the steering angle change,  $D$  is dataset size, and  $\delta t$  is the time between decisions.

### 101 3 Methods

102 **Baseline models** We adapt the PilotNet architecture’s output layer to produce a variety of baseline  
 103 models. The regression baseline has just one output node optimized to produce the steering value  
 104 via MAE loss. The classification baseline has the action space divided into  $N$  bins and optimizes  
 105 predicting the right bin via cross-entropy loss. Mixture density networks output between 1 to 5 triplets  
 106 of mean, standard deviation, and relevancy scores  $\alpha_i$  ( $\alpha_i \geq 0$ ,  $\sum_i \alpha_i = 1$ ), a linear combination of  
 107 which produces a Gaussian mixture model over action values. We use the mean of the most likely  
 108 Gaussian during deployment.

109 **EBM Training and Inference** We adapt the training and inference algorithms from energy-based  
 110 model literature [14] with a few modifications (see Figure 3). As a first modification, we use a  
 111 constant grid of linearly-spaced steering angles during training and inference instead of sampling  
 112 uniformly. Second, we do not use inference-time optimization to improve the initial candidate actions.  
 113 The candidate action values cover the steering angle range densely, and further optimization yielded  
 114 no gains. Furthermore, a fixed set of values is required by one of the proposed EBM modifications  
 115 and helps to make a cleaner comparison with the classification baseline. Early offline experiments  
 116 (see Appendix Figure 7) showed that these changes resulted in at least as good performance on  
 117 steering prediction as random sampling and inference-time optimization.

118 This results in the following loss function:

$$L_{EBM} = \text{CE}(\text{softmax}(-\mathbf{e}), \mathbf{y}) = - \sum_{i=1}^{n+1} (y_i \cdot \log(-\frac{e^{e_i}}{\sum_{j=1}^{n+1} e^{e_j}})), \quad (3)$$

119 where  $\mathbf{e}$  is the vector of energy values produced by neural network outputs  $e_i = E(\mathbf{o}, a_i)$  and  $\mathbf{y}$  is  
 120 one-hot vector having 1 at the position of ground truth action. Both  $\mathbf{e}$  and  $\mathbf{y}$  contain  $n + 1$  elements:  
 121  $n$  sampled values and one ground truth. To get the action energy vector  $\mathbf{e}$  for a single observation,  
 122 the neural network  $E$  is run on a batch of samples with the same repeated observation  $\mathbf{o}$  and different

123 actions  $a_i$ . In practice, the convolutional part of the network runs on an image only once, with actions  
124 fused into the model before the MLP head to reduce the time and memory usage.

125 Our initial implementation of EBM demonstrated a high whiteness score, i.e., high lateral jerk. This  
126 characteristic did not depend on the amount of training data (see Appendix Figure 6). We explored  
127 two changes to the EBM to combat this undesired characteristic.

128 **EBM with Temporal Smoothing** If one could reduce a model’s sensitivity to slight differences in  
129 subsequent camera frames, one would achieve temporally smoother predictions. An obvious choice in  
130 the case of energy-based models is to minimize the difference between predicted energy distributions  
131 at subsequent frames. So, we propose adding a temporal smoothness loss term, defined as:

$$L_{temp} = \alpha \| \mathbf{e}_t - \mathbf{e}_{t+1} \|, \quad (4)$$

132 where  $e_t$  stands for the vector of predicted energy values at timestep  $t$ , and  $\alpha$  is the smoothing strength.  
133 EBMs take actions as input, so  $e_t$  and  $e_{t+1}$  have to be computed with the same steering angle inputs,  
134 which motivated our use of a constant action grid instead of random sampling. However, to stick with  
135 the conventional sampling, one could also draw a random sample *once per pair* of frames. Since the  
136 ground truth steering angle is often different for consecutive frames, its energy values are masked  
137 from this loss term. A range of well-performing smoothing strengths was found empirically. We use  
138  $\alpha = 1.0$  for the temporally smoothed EBM in the final experiments.

139 **EBM with Soft Targets** We hypothesized that using one-hot targets in the cross-entropy loss  
140 is a major contribution to the higher whiteness of EBMs. Forcing nearby steering values to have  
141 drastically different energy is likely to make learning less efficient and the energy landscape noisier.  
142 This can lead to higher variance when choosing the best action via *argmin*.

143 Hence, we investigated a simple fix: use soft targets for the cross-entropy loss. Whereas soft targets  
144 have been widely used in neural networks with the purpose of regularization and better calibration  
145 [25, 23], our use case is a bit different. Unlike usual classification targets, our outputs are ordered,  
146 and we aim to enforce spatial smoothness. We replace the one-hot ground-truth vector with a vector  
147 assigning some of the probability to actions a few degrees away from the ground truth (see Figure 4).  
148 Target probabilities are computed as:

$$\mathbf{p}^* = \text{softmax}\left(\frac{(\mathbf{a} - a_{GT})^2}{T}\right), \quad (5)$$

149 where  $\mathbf{p}^*$  is a vector of target probabilities for cross-entropy loss,  $\mathbf{a}$  is the vector of input candidate  
150 steering values,  $a_{GT}$  is the ground truth steering angle, and  $T$  is the softmax temperature ( $2.5 * 10^{-3}$   
151 in all reported tests with soft targets). We picked the temperature value such that 99.9% of the  
152 probability mass was on  $\pm 5$  degrees around the ground truth.

## 153 4 Experimental Setup

154 **Dataset and Training Pipeline** We use the training dataset by Tampuu et al. [28], which consists  
155 of 540 km of human driving on WRC Rally Estonia tracks. These are usually very low-traffic gravel  
156 roads. The recordings that make up the dataset are broken into training (460 km) and evaluation  
157 (80 km), with the evaluation recordings used for off-policy metric calculation and early stopping.  
158 The original dataset includes camera and LiDAR images, but only camera frames are used in our  
159 experiments. Image pre-processing and training details are specified in the supplementary materials.

160 **On-Policy Evaluation** The on-policy evaluation was performed on a 4.3 km section of the WRC  
161 Estonia 2022 SS10+14 Elva track <sup>1</sup>, driven in both directions. No recordings from this track were in

<sup>1</sup><https://www.rally-maps.com/Rally-Estonia-2022/Elva>

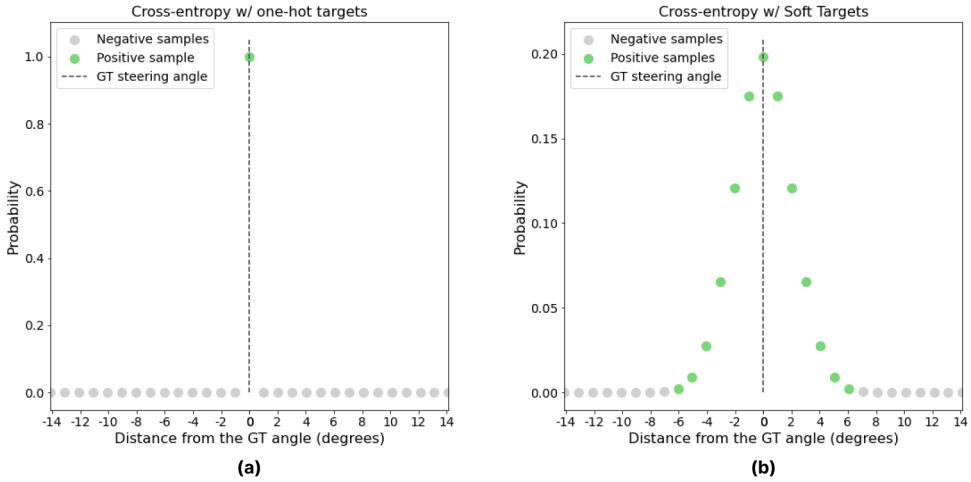


Figure 4: (a) One-hot (standard) cross-entropy pushes the probability of negative samples down, even if they are indistinguishably close to the target. (b) Soft targets give positive weights to the samples around the target proportionally to the distance. Note that the horizontal axis is trimmed to focus on the GT value; the full range of input actions we use is from  $\sim 250$  to  $\sim 250$  degrees.

162 the training set of models. The speed was set to 80% of the speed a human driver used in the same  
 163 location and direction, extracted from a prior recording. In practice, this meant a speed of up to 40  
 164 km/h. Setting the speed to 100% was attempted but felt too dangerous with certain models. The  
 165 testing was performed over multiple weeks of September 2022.

166 The evaluation track is narrow, and driving off the road edge is hazardous for the car, so the safety  
 167 driver was free to intervene when they perceived danger. An intervention was counted when the driver  
 168 applied force to turn the steering wheel. If the model turned the steering wheel simultaneously in the  
 169 same direction as the safety driver, it would not cause an intervention since no force was applied.

170 Alongside the intervention count, we also report the whiteness of steering as an on-policy metric.  
 171 Here, *command whiteness*  $W_{cmd}$  stands for the whiteness of the predicted steering commands during  
 172 on-policy evaluation, and *effective whiteness*  $W_{eff}$  refers to the resulting actual whiteness of the  
 173 front wheel angles as measured by the sensors. Command whiteness is usually higher than effective  
 174 whiteness due to the smoothing effect of the real-car actuators. No matter the force (i.e., the angular  
 175 acceleration of the steering wheel), it takes time to reach the target value.

176 First, around twenty test runs were completed to select the best representative of each model type  
 177 over several hyperparameters and random seeds. For the final experiments, the six most promising  
 178 models were chosen: an EBM with 512 linearly-spaced candidate values (standard, with temporal  
 179 smoothing, or with soft targets), a classification model with 512 bins, MDN with 5 Gaussians, and a  
 180 regression model with MAE loss. We performed four evaluation runs per model across four days.  
 181 The worst run for each model was discarded to account for out-of-distribution weather (excessive sun  
 182 or rain drops on camera) or safety driver variance.

183 **Evaluation in VISTA** Evaluating driving models in the real world can make it harder for other  
 184 researchers to replicate results. To aid reproducibility, we additionally run the main experiments in the  
 185 VISTA Driving Simulator[2]. VISTA is a data-driven simulator that allows replaying recordings of  
 186 real-world drives *interactively* by reprojecting the viewpoint as desired. Thus, a simulator can be used  
 187 for on-policy, closed-loop evaluation, allowing fast and reproducible model evaluation (as in standard  
 188 model-based simulators) while staying visually close to the real-world data distribution. To this end,  
 189 we release our evaluation code and the recording we used for evaluation in the VISTA format.<sup>2</sup> We

<sup>2</sup>Will be available in the de-anonymized version.

190 used a recording produced by the strongest model completing the track without interventions. This  
 191 recording had the highest correlation with our results on all models (see Appendix Table 3), due to  
 192 being most in-distribution with the weather and vegetation at the time of the real-world tests.

193 Absent a safety driver, we define crashes in VISTA as moments when the car drives more than 2  
 194 meters away from the expert-driven trajectory. After a crash, we restart the car two seconds further  
 195 down the road. This evaluation scheme has obvious limitations, for example, as the expert does not  
 196 always drive in the center of the road and 2 meters would be too much or too little to cause the safety  
 197 driver to disengage in reality. Yet, empirical results from the evaluation in VISTA support the findings  
 198 from our real-world experiments, suggesting that VISTA can be used to reproduce our key results.

## 199 5 Results

Table 1: Generalization results, with three real-world and three virtual driving sessions per model.

Model	Real world			VISTA	
	Interventions	$W_{eff}$	$W_{cmd}$	Crashes	$W_{cmd}$
EBM	4	35.25°/s	176.93°/s	2	114.33°/s
	1	32.34°/s	96.94°/s	1	121.57°/s
	2	28.57°/s	223.59°/s	2	121.67°/s
	mean:	2.33	32.05°/s	1.67	119.19°/s
EBM Temp. Smoothing	5	49.92°/s	119.39°/s	3	58.70°/s
	2	38.96°/s	137.22°/s	2	60.37°/s
	3	34.21°/s	77.28°/s	2	48.86°/s
	mean:	3.33	41.03°/s	1.33	55.98°/s
EBM Soft Targets	5	27.80°/s	56.33°/s	3	85.72°/s
	5	46.83°/s	57.15°/s	3	74.97°/s
	4	33.72°/s	56.86°/s	3	81.87°/s
	mean:	4.66	36.12°/s	1.33	80.85°/s
Regression (MAE)	2	26.69°/s	37.84°/s	0	24.39°/s
	2	29.65°/s	75.34°/s	0	24.75°/s
	1	26.28°/s	33.10°/s	0	24.25°/s
	mean:	1.66	27.54°/s	0	24.47°/s
Classification	1	41.05°/s	182.39°/s	1	123.69°/s
	7	62.17°/s	287.14°/s	1	105.13°/s
	1	34.11°/s	162.27°/s	1	104.31°/s
	mean:	3.00	45.77°/s	1	111.04°/s
MDN	1	25.32°/s	33.62°/s	3	37.22°/s
	5	24.82°/s	35.46°/s	3	35.74°/s
	5	26.66°/s	37.39°/s	3	35.84°/s
	mean:	3.66	25.59°/s	1.33	36.27°/s

200 The results of the final test runs are presented in Table 1. Each row in the table corresponds to  
 201 approximately 20 minutes of driving, so each model’s total intervention count is produced by an  
 202 hour of driving. We believe this is enough time to reveal noteworthy differences in performance on  
 203 a simple task such as road following. Real-world results are supported by VISTA showing similar  
 204 trends with lower variance (Pearson R = 87.5%, Spearman R = 92.7% for real-world interventions in  
 205 final experiments vs VISTA crashes).

206 We observe no clear benefit of using energy-based implicit behavioral cloning models over the explicit  
 207 baseline models. The baseline regression model resulted in the least interventions and had also the

208 most smooth driving according to the whiteness measure. The other two explicit baseline models:  
 209 classification and mixture density networks, resulted in more interventions across the three runs than  
 210 the EBM. Supported by VISTA, the results reveal that no one solution stands out clearly, and hence  
 211 that using the EBM formulation did not significantly improve the performance.

212 To bring the whiteness values of EBMs closer to the values of explicit models, we implemented two  
 213 variations to the EBM. Both approaches significantly reduced the whiteness of the model predictions  
 214 during deployment ( $W_{cmd}$  in Table 1). However, this reduction did not translate into a reduction of  
 215 effective whiteness, while both of these approaches resulted in more safety driver interventions.

216 The increased temporal stability of modified EBM approaches compared to the naive EBM is also  
 217 visible in Figure 5. This illustration visualizes the outputs of different models computed off-policy on  
 218 a recording of the vehicle passing an intersection. Furthermore, we see the classifier model would  
 219 swerve to the left if given control of the car near the intersection.

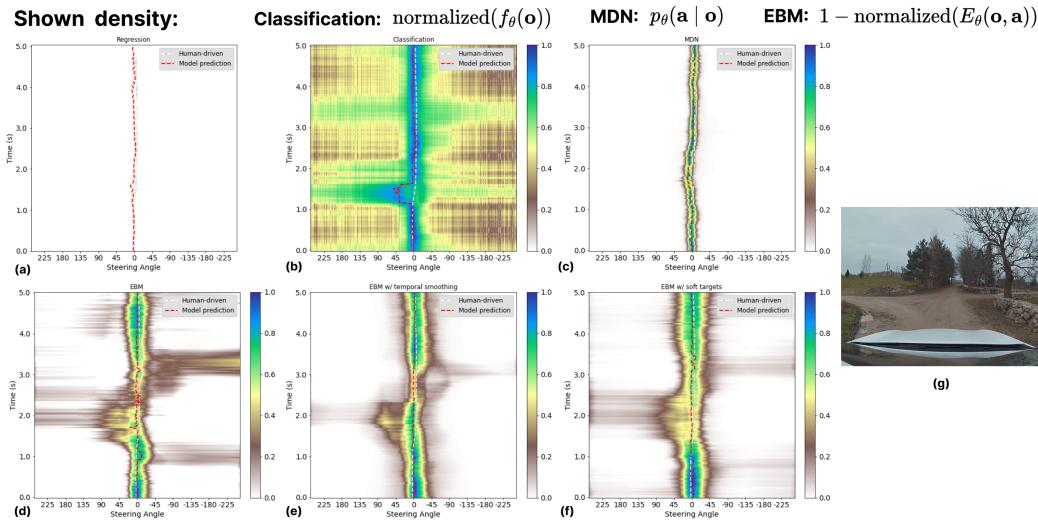


Figure 5: **Outputs of different models during 5 seconds of human driving (a-f).** The period corresponds to the car passing an intersection, with Y-axis representing the time and X-axis the predicted steering angle. The camera image 2 seconds into the recording is given on the right (g). The red dashed line represents the predicted steering values, grey dashed line is the ground truth.

220 EBM models exhibit slight multimodality  
 221 when passing the intersection with an area of  
 222 lower energy values towards the left, corre-  
 223 sponding to a left turn (Figure 5 (d)-(f)). The  
 224 standard EBM also shows an alternative hy-  
 225 pothesis of turning to the right around 3 sec-  
 226 onds into the recording. As reported above,  
 227 however, this ability to represent alternative  
 228 hypotheses on the energy landscape does not  
 229 translate into improved performance.

230 Supposedly, a model having an explicit rep-  
 231 resentation of alternative paths can select the  
 232 most likely and not produce intermediate be-  
 233 havior that is average of multiple options. At in-  
 234 tersections, such average behavior would show  
 235 as swerving towards the side road. When quan-  
 236 tified (see Table 2), classical regression models were shown to produce most swerving towards side

Table 2: **Handling multimodalities.** Swerve rate shows the percentage of challenging road sections where the model slightly swerved towards the side road. There are three such road sections, with three on-policy runs per model type. Slight swerves are difficult to set a threshold for and are counted as half a swerve. A lower number is better.

Model	Swerve rate
EBM	44%
EBM Temp. Smoothing	66%
EBM Soft Targets	39%
Regression	89%
Classification	44%
MDN	33%

237 roads. This observation aligns with the unimodal nature of regression models trained with the MAE  
238 loss. In contrast, other models swerved less, which can be attributed to their richer representations.

## 239 **6 Discussion**

240 This project investigated if the reported benefits of using the energy-based model formulation for  
241 behavioral cloning carry over to the task of real-world road following. We assumed that the claimed  
242 better generalization and handling of multimodalities could be useful in this task. However, the  
243 results show no improvements in overall driving ability. We hypothesize that our chosen task has too  
244 few multimodalities to make EBMs stand out. Prior work on implicit behavior cloning[14] used tasks  
245 where actions are distributed less normally than in road-following steering control (see Appendix  
246 Figure 8). Past tasks also had at least two-dimensional action spaces, which can make them more in  
247 need of multimodal policies. However, our models only predicted steering; and representing different  
248 hypotheses of steering angle was useful only at a few intersections along the route.

249 We did observe some improvement in handling situations that presumably require modeling multi-  
250 modal distributions, such as intersections and ignoring side roads. The only unimodal baseline,  
251 MAE-based regression, swerved towards side roads more frequently than other models. However, less  
252 frequent swerving of multimodal models did not result in fewer interventions overall. We attribute  
253 this to the low proportion of the task requiring a multimodal policy. Conversely, the unimodal loss  
254 seems to introduce an inductive bias, increasing the data efficiency in learning to handle simpler  
255 (unimodal) road situations that dominate the task.

256 We observed higher lateral jerk for multimodality-representing models. Our proposed modifications to  
257 the EBM training process significantly reduced the jerk of the model-predicted commands. However,  
258 this did not improve the effective whiteness of the car’s front wheels. We attribute this to the car’s  
259 actuators acting as a low-pass filter on the noise in the command sequence. This prevents even a large  
260 change in command whiteness from translating to a drop in front-wheel whiteness, motivating research  
261 on more powerful smoothing techniques. The temporal loss term was computed on consecutive  
262 images only 33 ms apart. However, effective whiteness seems to be caused by output variability on a  
263 slightly higher time scale. In future work, smoothing actions across a slightly longer timescale should  
264 be attempted. Soft targets proved surprisingly effective in reducing command-sequence whiteness,  
265 given that they do not directly enforce similarity across time.

## 266 **7 Conclusion**

267 We tested implicit behavioral cloning with energy-based models for controlling the steering of a  
268 real self-driving car. We showed that energy-based models perform comparably to classical explicit  
269 behavioral cloning baselines in terms of safety driver interventions but have higher jerk that reduces  
270 the comfort of the drive. We show two methods for reducing the steering jerk, measured as a whiteness  
271 score. Even though these methods greatly reduce the whiteness of predicted steering angles, it does  
272 not translate into improved whiteness of real steering, as the actuator delays in a real car smooth out  
273 radical steering movements anyway.

274 In our experiments, the simple regression-based explicit behavioral cloning baseline was the best  
275 in terms of interventions and jerk of the drive. However, the regression approach tended to swerve  
276 towards side roads, which comes from the unimodal nature of its loss function. We show that  
277 multimodality-capable models handle the situation with side roads better and do fewer swerves  
278 but do not eliminate the problem. Based on the analysis of action distributions in our task and in  
279 prior work where EBMs outperformed explicit models, we conclude that the lateral control in the  
280 road-following task has too few multimodalities to make EBMs useful. Altogether, this shows that  
281 while energy-based models have a number of theoretical advantages, it can be challenging to bring  
282 those out in real-world scenarios, and more research is needed to make efficient use of them.

283 **References**

- 284 [1] Carla leaderboard - evaluation and metrics, 2022. URL <https://leaderboard.carla.org/#evaluation-and-metrics>. 2
- 285
- 286 [2] A. Amini, T.-H. Wang, I. Gilitschenski, W. Schwarting, Z. Liu, S. Han, S. Karaman, and D. Rus. Vista 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022. 1, 4
- 287
- 288 [3] A. Bewley, J. Rigley, Y. Liu, J. Hawke, R. Shen, V.-D. Lam, and A. Kendall. Learning to drive from simulation without real world labels. In *2019 International conference on robotics and automation (ICRA)*, pages 4818–4824. IEEE, 2019. 2
- 289
- 290 [4] C. M. Bishop. Mixture density networks. 1994. 1
- 291
- 292 [5] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016. 1, 2
- 293
- 294 [6] M. Bojarski, C. Chen, J. Daw, A. Değirmenci, J. Deri, B. Firner, B. Flepp, S. Gogri, J. Hong, L. Jackel, et al. The nvidia pilotnet experiments. *arXiv preprint arXiv:2010.08776*, 2020. 2
- 295
- 296 [7] F. Codevilla, A. M. López, V. Koltun, and A. Dosovitskiy. On offline evaluation of vision-based driving models. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 236–251, 2018. 2
- 297
- 298 [8] F. Codevilla, M. Miiller, A. López, V. Koltun, and A. Dosovitskiy. End-to-end driving via conditional imitation learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–9. IEEE, 2018. 2
- 299
- 300 [9] F. Codevilla, E. Santana, A. M. López, and A. Gaidon. Exploring the limitations of behavior cloning for 301 autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 302 9329–9338, 2019. 2, 2
- 303
- 304 [10] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun. CARLA: An open urban driving simulator. 305 *arXiv preprint arXiv:1711.03938*, 2017. 2
- 306
- 307 [11] M. Elbanhawi, M. Simic, and R. Jazar. In the passenger seat: investigating ride comfort measures in 309 autonomous cars. *IEEE Intelligent transportation systems magazine*, 7(3):4–17, 2015. 2
- 310
- 311 [12] H. M. Eraqi, M. N. Moustafa, and J. Honer. End-to-end deep learning for steering autonomous vehicles 313 considering temporal dependencies, 2017. URL <https://arxiv.org/abs/1710.03804>. 2
- 312
- 313 [13] N. Fernandez. Two-stream convolutional networks for end-to-end learning of self-driving cars. 2018. doi: 10.48550/ARXIV.1811.05785. URL <https://arxiv.org/abs/1811.05785>. 2
- 314
- 315 [14] P. Florence, C. Lynch, A. Zeng, O. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mordatch, and J. Tompson. Implicit behavioral cloning. *Conference on Robot Learning (CoRL)*, 2021. 1, 2, 2, 3, 6, 8
- 316
- 317 [15] J. Hawke, R. Shen, C. Gurau, S. Sharma, D. Reda, N. Nikolov, P. Mazur, S. Micklenthwaite, N. Griffiths, 318 A. Shah, et al. Urban driving with conditional imitation learning. In *2020 IEEE International Conference 319 on Robotics and Automation (ICRA)*, pages 251–257. IEEE, 2020. 2
- 320
- 321 [16] J. Hawke, V. Badrinarayanan, A. Kendall, et al. Reimagining an autonomous vehicle. *arXiv preprint 322 arXiv:2108.05805*, 2021. 2
- 323
- 324 [17] S. Hecker, D. Dai, A. Liniger, M. Hahner, and L. Van Gool. Learning accurate and human-like driving 325 using semantic maps and attention. In *2020 IEEE/RSJ International Conference on Intelligent Robots and 326 Systems (IROS)*, pages 2346–2353. IEEE, 2020. 2
- 327
- 328 [18] A. Jain, L. Del Pero, H. Grimmett, and P. Ondruska. Autonomy 2.0: Why is self-driving always 5 years 329 away? *arXiv preprint arXiv:2107.08142*, 2021. 2
- 330
- 331 [19] N. Kalra and S. M. Paddock. Driving to safety: How many miles of driving would it take to demonstrate 332 autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, 94:182–193, 2016. 333 2
- 334
- 335 [20] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014. URL <https://arxiv.org/abs/1412.6980>. A.1
- 336
- 337
- 338
- 339
- 340

- 331 [21] A. Krogh and J. A. Hertz. A simple weight decay can improve generalization. In *Proceedings of the*  
332 *4th International Conference on Neural Information Processing Systems*, NIPS'91, page 950–957, San  
333 Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc. ISBN 1558602224. A.1
- 334 [22] Y. Lecun, S. Chopra, and R. Hadsell. *A tutorial on energy-based learning*. 01 2006. 1, 2
- 335 [23] R. Müller, S. Kornblith, and G. E. Hinton. When does label smoothing help? *Advances in neural*  
336 *information processing systems*, 32, 2019. 3
- 337 [24] B. Nassi, D. Nassi, R. Ben-Netanel, Y. Mirsky, O. Drokin, and Y. Elovici. Phantom of the adas: Phantom  
338 attacks on driver-assistance systems. <https://eprint.iacr.org/2020/085.pdf>, 2020. 2
- 339 [25] G. Pereyra, G. Tucker, J. Chorowski, Ł. Kaiser, and G. Hinton. Regularizing neural networks by penalizing  
340 confident output distributions. *arXiv preprint arXiv:1701.06548*, 2017. 3
- 341 [26] D. A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In *Advances in neural*  
342 *information processing systems*, pages 305–313, 1989. 2
- 343 [27] A. Tampuu, T. Matiisen, M. Semikin, D. Fishman, and N. Muhammad. A survey of end-to-end driving:  
344 Architectures and training methods. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.  
345 1, 2, 2
- 346 [28] A. Tampuu, R. Aidla, J. A. van Gent, and T. Matiisen. Lidar-as-camera for end-to-end driving, 2022. URL  
347 <https://arxiv.org/abs/2206.15170>. 2, 4
- 348 [29] W. Zeng, W. Luo, S. Suo, A. Sadat, B. Yang, S. Casas, and R. Urtasun. End-to-end interpretable neural  
349 motion planner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*,  
350 pages 8660–8669, 2019. 2

351 **A Appendix**

352 **A.1 Pre-processing, training, and hardware**

353 **Pre-processing** The frames are cropped to remove the car’s hood and everything beyond the horizon  
 354 and to limit the view to 90 degrees of the front center. The resulting frames of shape 264x68x3 are  
 355 then normalized and fed into the model. The target labels correspond to the steering wheel angles of  
 356 the human drivers.

357 **Training** When training, a mini-batch is created by sampling uniformly from all recordings. For  
 358 experiments with temporal smoothing, a different sampling approach is used, where sequences of two  
 359 consecutive frames are sampled instead. The sequence dimension is flattened such that a mini-batch  
 360 of sequences becomes a mini-batch of frames.

361 The Adam [20] optimizer is used with default hyperparameters (learning rate  $1 * 10^{-3}$ , betas 0.9 and  
 362 0.999) and  $1 * 10^{-2}$  weight decay [21]. Finally, early stopping is used on validation MAE with a  
 363 patience of 10 epochs.

364 **Car Hardware and Software Stack** We perform the experiments with Lexus RX 450h fitted with  
 365 a PACMod v3 drive-by-wire system. The following sensors are used: a NovAtel PwrPak7D-E2  
 366 GNSS device and a Sekonix SF3324 120-degree FOV camera. The car computer is equipped with a  
 367 GeForce GTX 2080 GPU. The camera works at 30 Hz, but our end-to-end stack is slower ( $\sim$ 12 Hz).  
 368 To accommodate for the differing processing speeds all but the latest frame in the queue are dropped.

369 **A.2 Validation loss and whiteness with different data amounts**

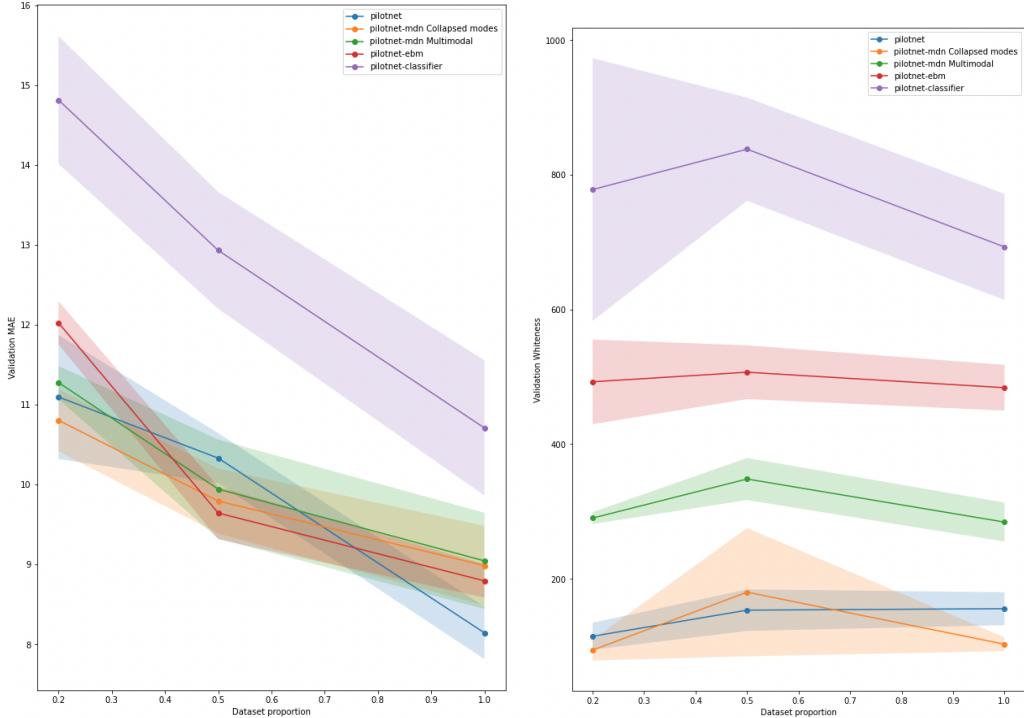


Figure 6: **Left:** Varying the dataset size predictably changes the predictive accuracy of all model formulations. **Right:** Whiteness does not seem to be influenced by the amount of data when the amount was varied by 5 times.

370 **A.3 How do our EBM simplifications affect modeling performance?**

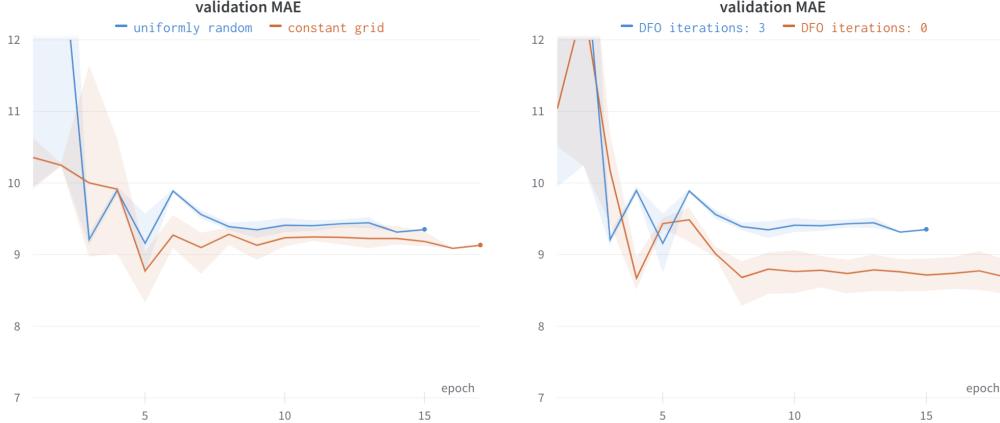


Figure 7: **Left:** Using a constant grid of actions results in at least as good MAE as the more common uniformly random sample on each decision. **Right:** Derivative-free optimization does not improve performance beyond a one-shot argmin; to our surprise, it even hurts validation MAE.

371 **A.4 Is road following just too unimodal?**

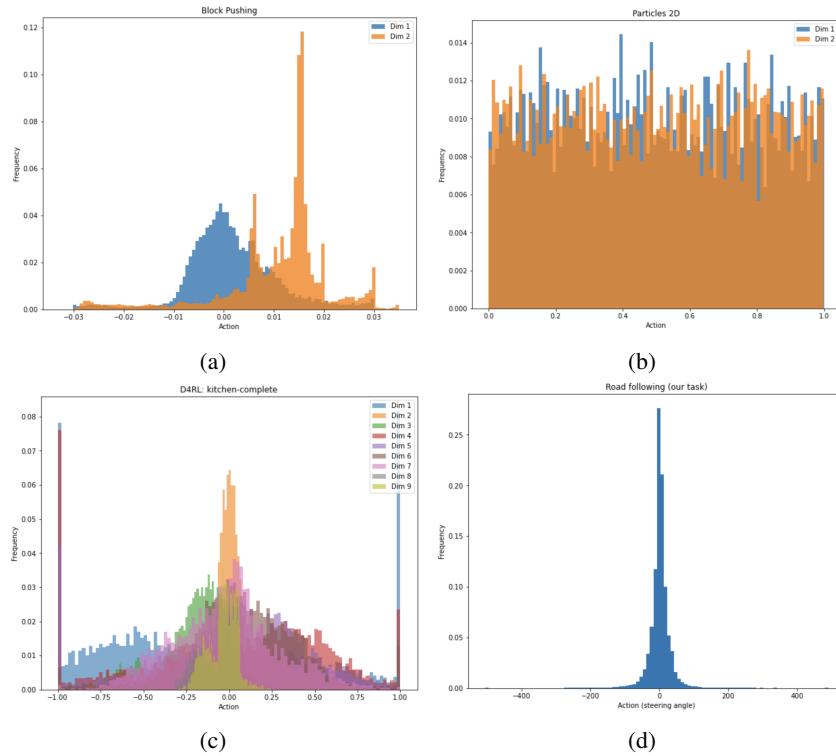


Figure 8: Action distributions from three randomly-picked example tasks from prior work on IBC[14] where EBMs outperformed explicit models (a-c). Action distribution for our road-following task (d) looks much more gaussian, which is a hint for a lower number of possible multimodalities.

372 **A.5 VISTA agreement with real-world results**

373 In the main text, we report VISTA’s agreement with the results in the main experiments. In Table 3,  
374 we report the results for all models we tested throughout the project. When run on the same track,  
375 VISTA seems to have a very high agreement with real-world results.

Table 3: Per-model mean metrics correlations for VISTA vs reality (n=17 models).

Measure	Interventions	$W_{cmd}$
Pearson	83%	89%
Spearman	84%	86%