

The Importance of Balanced Data Sets: Analyzing a Vehicle Trajectory Prediction Model based on Neural Networks and Distributed Representations

1st Florian Mirus
Research, New Technologies, Innovations
BMW AG
Garching, Germany
florian.mirus@bmwgroup.com

2nd Terrence C. Stewart
Applied Brain Research Inc.
Waterloo, Ontario, Canada
terry.stewart@appliedbrainresearch.com

3rd Jörg Conradt
Dep. of Comp. Science and Technology
KTH Royal Institute of Technology
Stockholm, Sweden
conr@kth.se

Abstract—Predicting future behavior of other traffic participants is an essential task that needs to be solved by automated vehicles and human drivers alike to achieve safe and situation-aware driving. Modern approaches to vehicles trajectory prediction typically rely on data-driven models like neural networks, in particular LSTMs (Long Short-Term Memorys), achieving promising results. However, the question of optimal composition of the underlying training data has received less attention. In this paper, we expand on previous work on vehicle trajectory prediction based on neural network models employing distributed representations to encode automotive scenes in a semantic vector substrate. We analyze the influence of variations in the training data on the performance of our prediction models. Thereby, we show that the models employing our semantic vector representation outperform the numerical model when trained on an adequate data set and thereby, that the composition of training data in vehicle trajectory prediction is crucial for successful training. We conduct our analysis on challenging real-world driving data.

Index Terms—Vehicle trajectory prediction, neural networks, LSTMs, distributed representations, data set composition

I. INTRODUCTION

The race to autonomous driving is currently one of the main forces for pushing research forward in the automotive domain. With highly automated vehicle prototypes gradually making their way to our public roads and fully-automated driving on the horizon, it seems to be a matter of time until we see robot taxis or cars navigating us through urban traffic or heavy stop-and-go on highways. One major reason for this development in recent years is the rapid progress of AI (Artificial Intelligence), especially the success of deep learning, which has shown remarkable results in tasks essential for automated driving like object detection, classification [1] and control [2].

Predicting future behavior and positions of other traffic participants from observations is a cornerstone for successful collision avoidance and thus safe motion planning, and needs to be solved by human drivers and automated vehicles alike to reach their desired goal. Motion prediction for intelligent vehicles in general has seen extensive research in recent years [3]–[6]. Lefèvre et al. [4] classify such prediction approaches into three categories, namely *physics-based*, *maneuver-based*

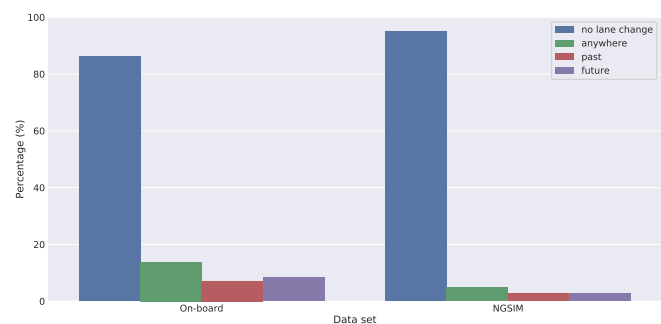


Fig. 1. Visualization of the composition of both our data sets regarding lane changes of the target vehicle.

and *interaction-aware*, depending on their level of abstraction. *Physics-based* and *maneuver-based* motion models consider the law of physics and the intended driving maneuver respectively as the only influencing factors for future vehicle motion and ignore inter-dependencies between the motion of different vehicles. There exists a growing number of different *interaction-aware* approaches to account for those dependencies and mutual influences between traffic participants or, more generally, agents in the scene. Probabilistic models like costmaps [7] account for physical constraints on the movements of the other vehicles. Classification approaches categorize and represent scenes in a hierarchy [8] based on the most generic ones to predict behavior for a variety of different situations.

Data-driven approaches to behavior prediction mainly rely on LSTM (Long Short-Term Memory) neural network architectures [9], which have proven to be a powerful tool for sequential data analysis. Alahi et al. [10] model interactions in the learning network architecture by introducing so-called social-pooling layers to connect several LSTMs each predicting the distribution of the trajectory position of one agent at a time. Deo and Trivedi [11] adapted the combination of LSTM networks for encoding vehicle trajectories and (convolutional) social-pooling layers to account for interactions to vehicle pre-

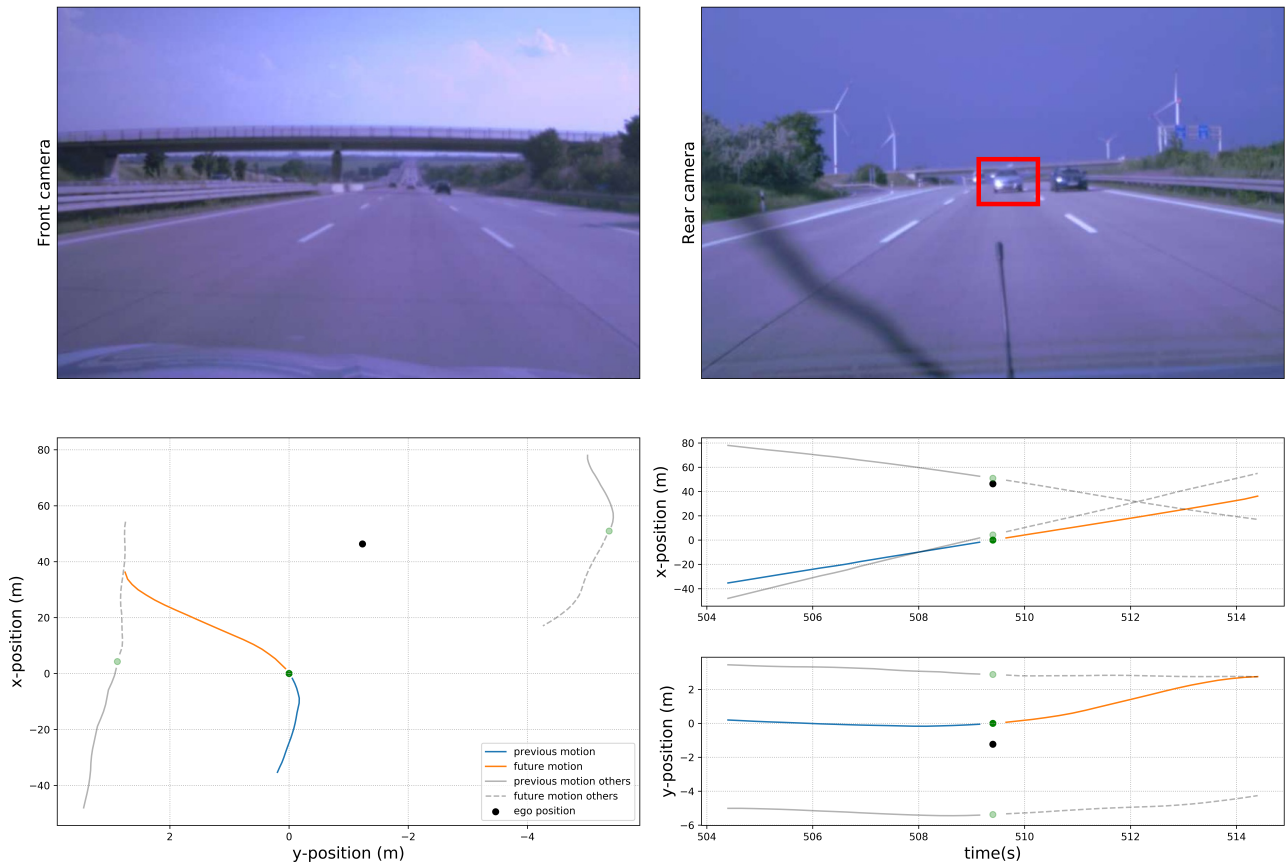


Fig. 2. Data visualization of one data sample from the *On-board* data set D_1 containing a future lane change of the target vehicle. The dots in the left plot indicate the position of the vehicles and color-code the vehicle type (red=motorcycle, green=car, blue=truck, black=ego-vehicle), blue and orange lines show past and future motion of the target vehicle whereas gray lines depict the other vehicles’ motion. The images in the top row show raw images recorded using the ego-vehicle’s front and rear camera with the target vehicle highlighted by a red bounding box.

diction in highway situations. Altche and de La Fortelle [12] use a LSTM network as well, but they account for interactions by including distances between the target vehicle and other agents directly in the training data rather than adapting the network architecture. A similar approach is proposed in [13], but it combines LSTM networks with an additional maneuver classification network to predict future vehicle motion.

However, for any model employing a data-driven learning approach, the underlying data set used for training is of crucial importance for the model to successfully learn and, more importantly, generalize beyond unseen examples. Compared to the field of computer vision with many publicly available data sets such as Imagenet [14], MS COCO (Microsoft Common Objects in Context) [15] or, specifically in automotive context, the KITTI [16], Cityscapes [17] or ApolloScapes [18] data sets for training and evaluating data-driven models, there is a lack of publicly available data sets for vehicle trajectory prediction. The currently most widely used data set for trajectory prediction [11]–[13], [19] is the *NGSIM* (*Next Generation Simulation*) data set [20], which was originally created for traffic flow analysis and contains only examples from highway driving. Furthermore, this data set suffers from a strong imbalance towards straight driving compared to more

“interesting” driving maneuvers such as lane changes (cf. Fig. 1). The recently proposed INTERACTION data set [21] as well as the trajectory data set from ApolloScapes [18] are a first step to tackle this issue.

A. Contribution

In this work, we adopt LSTM-based vehicle prediction models as well as an environment representation based on distributed representations from previous work [19]. We found that vehicle prediction models could benefit from our vector representation in certain situations such as, for instance, crowded driving situations with multiple vehicles in close surrounding of the target- and/or ego-vehicle. In this paper, we extend that work by analyzing the influence of the data set used for training on the models’ performance. Here, we focus on comparing models trained on the complete set of samples or solely on training samples containing a lane change performed by the target vehicle (cf. Fig. 2). Following this approach, our contribution is two-fold: firstly, we are able to show that models employing our semantic vector representation proposed in [19], which encodes information about vehicle interactions directly within the semantics of the vectors themselves, perform better than models using a

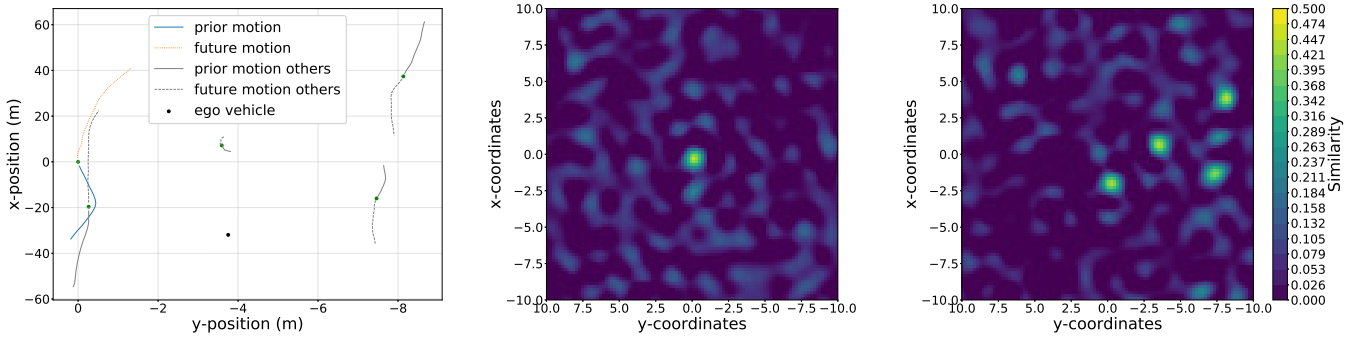


Fig. 3. Visualization of the convolutive power representation for 512-dimensional vectors. The left plot depicts a scene from the *On-board* data set, while the middle and right plots visualize the similarity between the representation vector of that scene and auxiliary comparison vectors created from a sequence of discrete values as heat map for the target vehicle (middle) and other cars (right).

simple numerical encoding of the input data without access to mutual interactions between agents when trained on lane change samples. Secondly, we find that the composition of the data used for training models for vehicle trajectory prediction is crucial for successful learning and should be considered in more detail in future research.

II. MATERIALS AND METHODS

A. Data and Preprocessing

In this work, we use two different data sets for training and evaluation of our system, which we refer to as *On-board* or D_1 , which is a proprietary data set containing real-world data gathered during test drives mainly on highways in southern Germany, and *NGSIM* or D_2 , which is a publicly available data set recorded using external cameras observing a segment of approximately 640 m length with 6 lanes on the US-101 freeway in Los Angeles, California. Both data sets contain object-lists with a variety of features such as position, velocity and acceleration but also object type probabilities and lane information. The *On-board* data set contains 3891 vehicles, which yield a total length of roughly 28.3 h when adding up the time each individual vehicle is visible, whereas the *NGSIM* data set contains 5930 vehicles and therefore a total time of roughly 91.3 h. For training and evaluating our model, we split both data sets into training $T_i \subset D_i$ and validation data $V_i \subset D_i$ containing 90 % and 10 % of the objects respectively to avoid testing our models on vehicles they have been trained with. Although we used both data sets in previous work [19], we conduct our analysis regarding the composition of the training data in this paper, however, solely on the *On-board* data set and only state a similar or even more imbalanced composition of the data set with respect to lane change maneuvers for the *NGSIM* data set (cf. Fig. 1)

B. Data set composition

In this section, we analyze the composition of our data sets regarding the amount of “interesting” behavior of the target vehicle. Both, the *On-board* and *NGSIM* data set consist of mainly highway driving, where we would expect mainly straight driving with the most interesting situations being the target vehicle, i.e., the vehicle whose motion we aim to

predict, performing a lane change. Hence, we are interested in the amount of situations where the target vehicle actually performs a lane change and how much more prominent normal straight driving is in both our data sets. For the *On-board* data set, we have information about the current lane as well as the distance to the lane borders estimated from the ego-vehicle’s cameras available for all vehicles. The *NGSIM* data set contains information about the current lane for each vehicle extracted from the external camera’s video footage. Thus, the selection process for the examples containing a lane change is straightforward for both data sets. Figure 2 shows one data sample from the *On-board* data set containing a lane change performed by the target vehicle in its future motion to be predicted. Comparing this example to the one shown in Fig. 3, which shows mainly straight driving for all vehicles present in the scene, we observe that a lane change mainly influences the vehicle’s motion in lateral (y) direction.

Figure 1 shows the amount of situations where the target vehicle performs a lane change in comparison to the amount of situations where no such behavior occurs for both, the *On-board* and *NGSIM* data set. For the *On-board* data set, in 86.1 % of all data samples the target vehicle does not perform a lane change, i.e., only 13.8 % of all data samples contain a lane change performed by the target vehicle. We further distinguish between lane changes performed during the trajectory history, i.e., the past 5 s before the current time step (labeled as *past* in Fig. 1) and lane changes that are performed in the future, i.e., during the future 5 s from the current time step (labeled as *future* in Fig. 1). We consider the lane changes in the future part of data samples to be the most interesting and challenging ones, since any model making predictions about the future trajectory needs to be able to anticipate these lane changes. For the *On-board* data set, 7 % of all data samples contain a lane change in the trajectory history, while 8.2 % of the samples contain a future lane change performed by the target vehicle. In comparison to the 86.1 % of data samples not containing a lane change, the amount of samples with interesting behavior other than straight driving within the data set is significantly less present. For the *NGSIM* data set, the discrepancy between the amount of samples without the target vehicle performing a lane change and the number of samples containing a lane change

is even more significant. The percentage of samples without a target vehicle lane change is 95.1% while only 4.9% of the samples contain a lane change performed by the target vehicle at all. The amount of samples containing a future lane change performed by the target vehicle is only 2.6% of all samples in the *NGSIM* data set. Hence, there is a significant imbalance in both data sets between examples containing mainly straight driving by the target vehicle, namely 86.1% and 95.1% of all samples in the *On-board* and *NGSIM* data set respectively, where most likely already simple prediction approaches are able to achieve reasonable results.

C. Convolutional vector-power

The SPA (Semantic Pointer Architecture) [22] is one special case of Vector Symbolic Architectures [23], a family of modeling approaches based on high-dimensional vector representations. Such vectors are one example of distributed representations in the sense that information is captured over all dimensions of the vector instead of one single number, which allows to encode both, symbol-like and numerical structures in a similar and unified way. Additionally, the architectures' algebraic operations allow manipulation and combination of represented entities into structured representations. In our work, atomic vectors are picked from the real-valued unit sphere, the dot product serves as a measure of similarity and the algebraic operations are component-wise vector addition \oplus and circular convolution \otimes . In this work, we make use of the fact that for any two vectors \mathbf{v}, \mathbf{w} , we can write

$$\mathbf{v} \otimes \mathbf{w} = \text{IDFT} \left(\text{DFT}(\mathbf{v}) \odot \text{DFT}(\mathbf{w}) \right), \quad (1)$$

where \odot denotes element-wise multiplication, DFT and IDFT denote the Discrete Fourier Transform and Inverse Discrete Fourier Transform respectively.

Using Eq. (1), we define the *convolutional power* of a vector \mathbf{v} by an exponent $p \in \mathbb{R}$ as

$$\mathbf{v}^p := \Re \left(\text{IDFT} \left((\text{DFT}_j(\mathbf{v})^p)_{j=0}^{D-1} \right) \right), \quad (2)$$

where \Re denotes the real part of a complex number. Furthermore, we call a vector \mathbf{u} *unitary*, if $\|\mathbf{v}\| = \|\mathbf{v} \otimes \mathbf{u}\|$ for any other \mathbf{v} (see [24, Sec. 3.6.3 and 3.6.5] for more details on the convolutional power and unitary vectors).

D. Vector representation

In this paper, we adopt the vector representation of automotive scenes for trajectory prediction [19] with the goal to evaluate the performance of the models proposed in [19] when trained and evaluated on different data set variations. We assign a random ID-vector to each category of dynamic objects (e.g., car, motorcycle, truck) as well as random unitary vectors \mathbf{X} and \mathbf{Y} to encode spatial positions. Let (x, y) denote the position of the target vehicle and (x_{obj}, y_{obj}) the positions of all other visible objects closer than 40m to the target (to

Short name	Input	Position encoding	Prediction model	Data set
linear	current position and velocity	-	Linear regression	both
LSTM numerical	sequence of positions	-	LSTM	both
LSTM SPA 1	semantic vector sequence	convolutional power	LSTM	both
LSTM SPA 2	semantic vector sequence	scalar multiplication	LSTM	both
LSTM SPA 3	semantic vector sequence	convolutional power incl. ego-vehicle	LSTM	On-board

TABLE I
SUMMARY OF THE EMPLOYED MODELS REGARDING LEARNING APPROACH, INPUT DATA AND POSITION ENCODING.

avoid accumulation of noise), we encapsulate this information in a scene vector

$$\mathbf{S} = \mathbf{TARGET} \otimes \mathbf{TYPE}_{target} \otimes \mathbf{X}^x \otimes \mathbf{Y}^y \oplus \sum_{obj} \mathbf{TYPE}_{obj} \otimes \mathbf{X}^{x_{obj}} \otimes \mathbf{Y}^{y_{obj}}, \quad (3)$$

where **TARGET** denotes an additional ID-vector chosen at random to indicate the target object (relevant for trajectory prediction). Figure 3 visualizes one example scene from the *On-board* data set and its representation vector queried for the target and other cars.

E. Prediction models

In this work, we analyze the LSTM (Long Short-Term Memory) models used in [19] for the prediction of vehicle positions regarding variations in the data used for training and evaluation. All networks consist of one LSTM encoder and decoder cell for sequence to sequence prediction, which means that the input and the final result of our model is sequential data. The encoder LSTM takes positional data for 20 past, equidistant time frames as input. That is, the input data is a sequence of 20 items of either positions of the target vehicle or a sequence of high-dimensional vectors encoding this positional data (see Sec. II-D and Table I for further details). Thus, the resulting embedding vector encodes the history of the input data over those 20 time frames. Finally, the embedding vector is used as input for the decoder LSTM to predict future vehicle positions. The output of each model is a sequence of 20 positions of the target vehicle predicted over a certain temporal horizon into the future. The only difference is the dimensionality and encoding of the input data. Similar to [19], we use a linear regression model based on a constant velocity assumption as a simple baseline for reference. Table I gives an overview of the models used in this paper. Further details on implementation and training of the individual models can be found in [19].

III. EXPERIMENTS

In section II-B, we have already seen that our prediction models using neural networks have to deal with imbalanced data sets containing significantly more straight driving than lane changes performed by the target vehicle. Hence, training

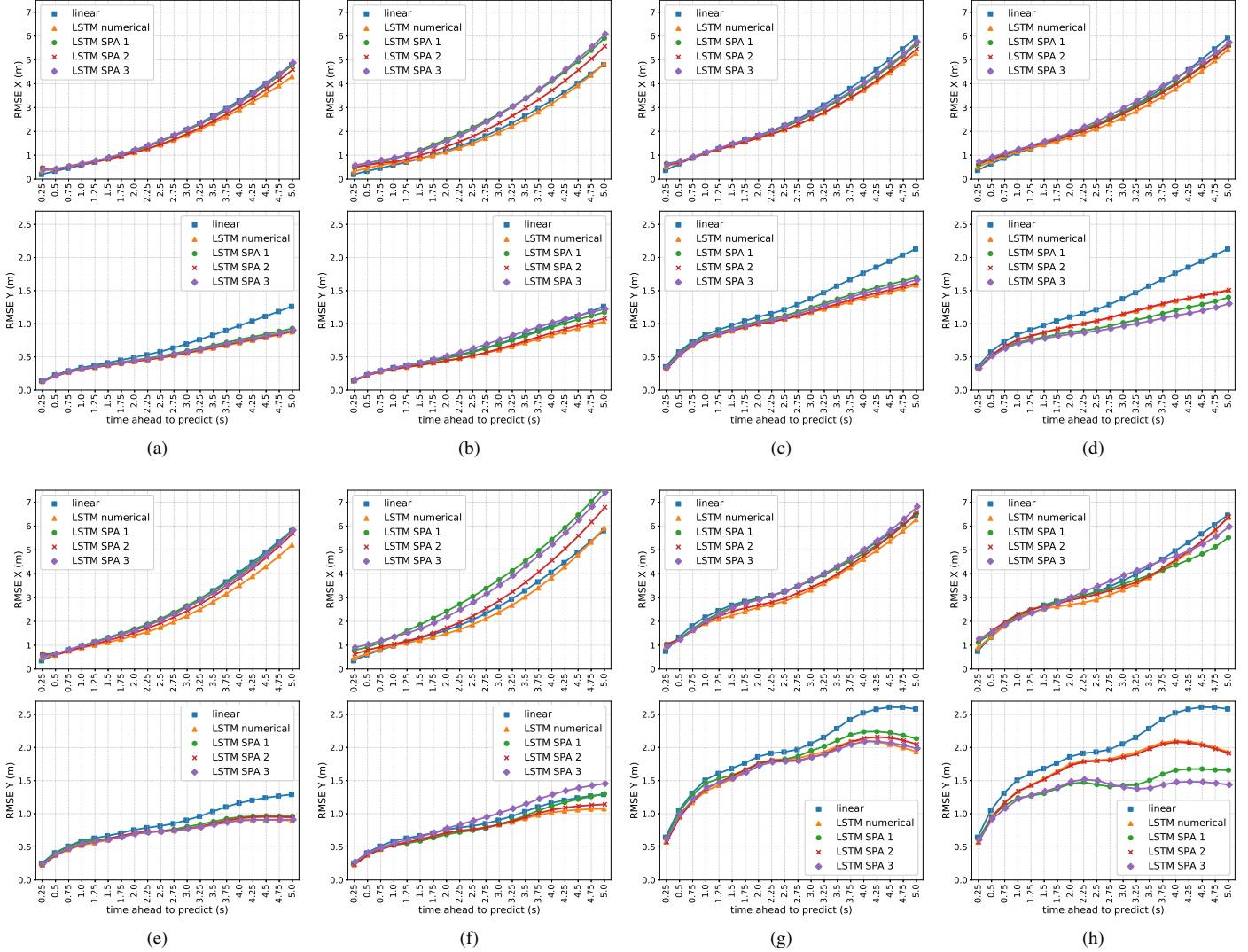


Fig. 4. Visualization of the RMSE performance of our LSTM models for different data setups. Figures (a), (c), (e) and (g) show the RMSE for models trained on the complete data set, while Fig. (b), (d), (f), (h) show the same models evaluated on the same samples but trained only on the samples including a target vehicle lane change. Table II gives a more detailed overview of the setup for each individual sub-figure.

Setup ID	Training set	Evaluation set
(a)	all samples	all samples
(b)	lane change samples	all samples
(c)	all samples	lane change samples
(d)	lane change samples	lane change samples
(e)	all samples	crowded samples
(f)	lane change samples	crowded samples
(g)	all samples	crowded lane change samples
(h)	lane change samples	crowded lane change samples

TABLE II
SUMMARY OF THE DATA SAMPLES USED FOR THE EVALUATIONS SHOWN IN INDIVIDUAL SUB-FIGURES OF FIG. 4.

any learning system on all the samples of both our data sets will expose the system to a significantly higher amount of data, in which most likely already a simple linear prediction model performs reasonably well. In this section, we therefore analyze the influence of the training data set on our LSTM models and if there are significant differences in the performance

of models trained on the complete data sets and on subsets consisting only of such samples that contain a lane change of the target vehicle. We conduct this analysis on the *On-board* data set only. Figure 4a shows the performance of our LSTM-based models trained on the complete data set. Here, we train models with the exact same neural network architectures and encoding schemes of the input just on different data sets, namely the subset of samples containing a lane change performed by the target vehicle. We consider both types of target vehicle lane changes, namely those performed during the trajectory history and lane changes performed in the future trajectory to be predicted, as input for the models. Figure 4 shows a comparison of different setups regarding training and evaluation data for our LSTM-based trajectory prediction models. Table II gives the details of the training and evaluation data setups visualized in Fig. 4. For the term “crowded

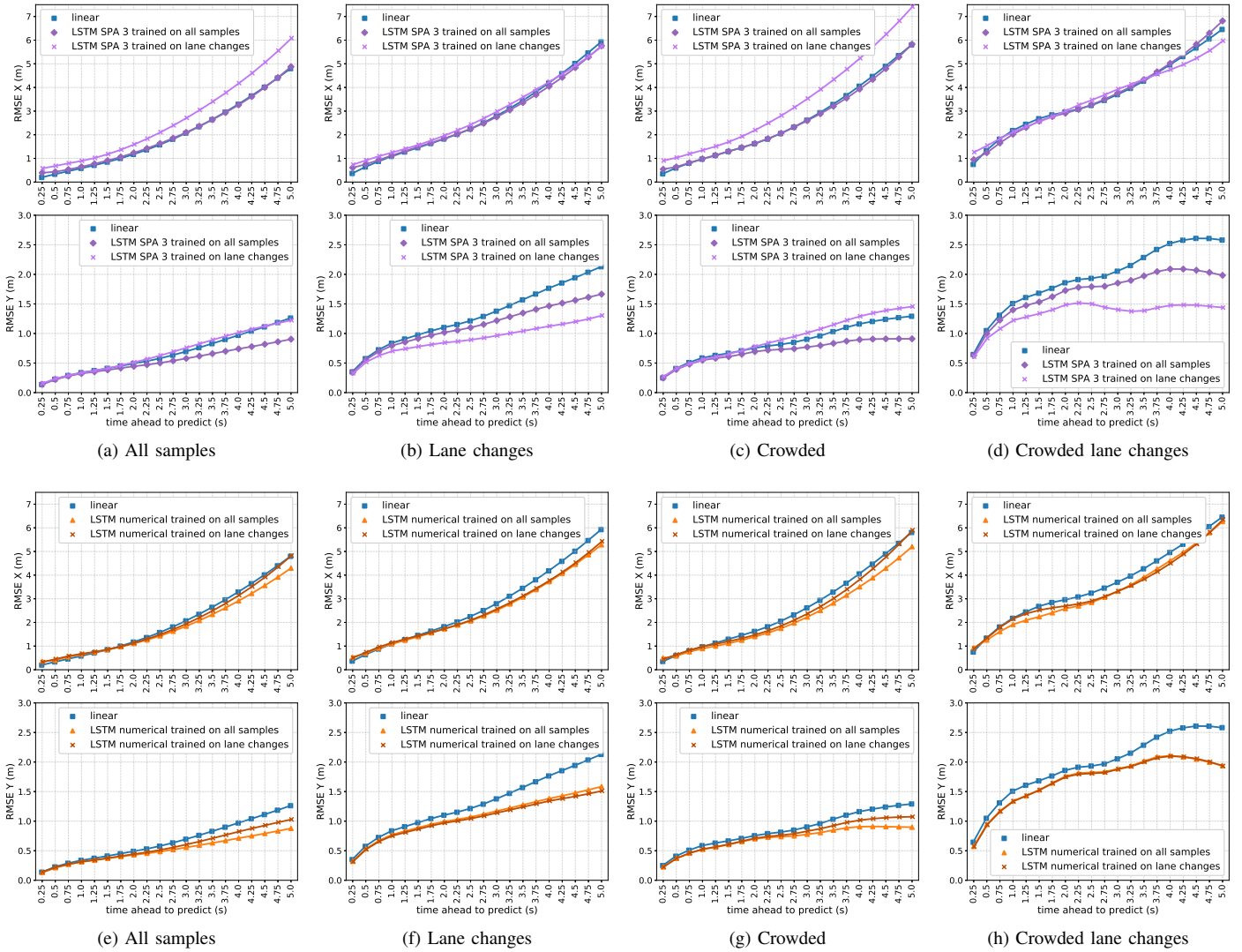


Fig. 5. Visualization of the changing RMSE performance of particular prediction models depending on the data they have been trained on. The first four figures (a) - (d) illustrate the difference between the LSTM SPA 3 models when changing their training data, while the last four figures (e) - (h) shows the same comparison for the LSTM numerical models.

samples” as mentioned in Table II, we refer to the metrics established in [19]: we consider samples with a distance less than 20 m between the target- and ego-vehicle, with distances less than 20 m between the target vehicle and the closest other vehicle, and with at least 3 other vehicles present in the scene as crowded situations. Furthermore, Table I gives an overview of the different prediction models and their short name used in the legends of Fig. 4 and 5. In both, Fig. 4 and 5, we also show a simple linear prediction model based on a constant velocity assumption for reference, since it is not a data-driven learning model and is therefore invariant under the changes of the training data. Figures 4a, 4c, 4e and 4g show the RMSE for models trained on the complete data set while Fig. 4b, 4d, 4f and 4h show the same models evaluated on the same samples but trained only on data samples including a target vehicle lane change. On the other hand, the upper row of Fig. 4, i.e.,

Fig. 4a - 4d illustrates the performance of the models trained on either the complete data set or on only the lane changes evaluated on all data samples, while the lower row, i.e., Fig. 4e - 4h shows a similar evaluation for crowded driving situations.

We observe that the models that have been trained only on samples containing a lane change performed by the target vehicle tend to achieve worse results than the models trained on the complete data set, when being evaluated on the entirety of all data samples (Fig. 4a, 4e, 4c and 4g). Interestingly, the performance of the models based on the convolutive power encoding scheme (LSTM SPA 1 and 3) deteriorates more significantly compared to the other data-driven models, especially in lateral (y) direction. However, if we evaluate the same models only on the samples containing a target vehicle lane change, their performance changes significantly (Fig. 4b, 4d, 4f and 4h). We recapitulate the findings of section II-B that lane changes

influence the lateral (y) position values more severely than the longitudinal (x) position compared to straight driving samples. Therefore, the performance difference between the models trained on lane changes only and models trained on the complete data set, as we would expect, is also not that significant in longitudinal direction. Considering the lateral (y) direction however, we observe a significant change between both model and evaluation variants. If the models are trained only on lane change samples, the LSTM SPA 1 and 3 models outperform all other models in lateral direction when evaluated only on the data samples containing a lane change while their performance in longitudinal direction does not change significantly compared to the models trained on the complete data set.

So far, we have only compared either all models trained on the complete data set or all models trained only on the lane change samples. However, we are also interested in how the performance of particular models changes when modifying the underlying training data set. Figure 5 shows a comparison of the LSTM-based models when modifying the training data for the LSTM SPA 3 model (Fig. 5a - 5d) as well as the LSTM numerical model (Fig. 5e - 5h). In this direct comparison, we observe that there is no significant difference in the performance of the numerical LSTM models trained on different samples for both evaluation sets containing either the complete data set or only the target vehicle lane changes. For the LSTM SPA 3 model however, we observe significant improvements for the model trained on the lane change samples when evaluated on the lane changes performed by the target vehicle compared to the model trained on all data samples. This result could indicate that the numerical model trained on all samples generalizes sufficiently well to all possible situations compared to the convolutive power based model. However, we have already seen, that both trajectory data sets show a significant imbalance towards straight driving compared to lane change maneuvers (cf. section II-B), which is the same for all models. Therefore, we believe that the results shown here rather suggest that the learning models employing the convolutive vector-power and thereby encapsulating the prior motion not only of the target vehicle but also other vehicles in its surroundings are better suited to predict lane change maneuvers given a more balanced data set.

IV. DISCUSSION

A. Conclusion

In this paper, we have analyzed the influence of the composition of training data sets on neural network based models for vehicle trajectory prediction. In particular, we focused on models employing LSTMs in combination with a semantic vector representation of automotive scenes and compared them with similar networks employing simple numerical representations of the input data. We found that when training the LSTM models only on data samples containing a lane change performed by the target vehicle, that the model employing the convolutive vector-power representation outperforms all other approaches in y -direction when evaluated on the samples

containing a lane change, especially in crowded and potentially dangerous situations. These results suggest that training the whole system on a more balanced data set containing equally many lane change and straight driving samples could improve overall model performance. Furthermore, the results suggest that the networks using semantic vectors benefit from a more focused and specific training data set, since the models trained particularly on lane changes improved especially in lane change situations compared to network variants trained on all samples. We expect that similar observations regarding data set imbalance as described in this paper also occur when replacing LSTMs with other neural network architectures for sequence to sequence prediction such as TCNs (Temporal Convolutional Networks) [25], but we leave a more detailed investigation of other network architectures for future work. Finally, there is not only room for improvement for the models investigated here, but also other data-driven models taking interactions between several vehicles into account used for trajectory prediction in general, by researching and evaluating the influence of the distribution of driving situations within the training data.

B. Future work

The results presented in this paper indicate that distributed representations are an interesting option to encode automotive scenes, especially when modeling mutual interactions between traffic participants and using the representational substrate as input to neural networks. However, one key strength of such representations and therefore essential for them to unfold their full potential, is the possibility of implementation in SNNs (Spiking Neural Networks) [22] and therefore deployment on dedicated neuromorphic hardware. This deployment can only be achieved in combination with prediction models employing a spiking neuron substrate as well, which could be an interesting, energy-efficient option in future automated vehicles with tight restrictions regarding on-board power consumption. Unfortunately, LSTM-based neural networks neither allow implementation in spiking neurons nor deployment on neuromorphic hardware. Therefore, one direction for future research could be to replace the LSTMs in our current approach with the recently proposed LMUs (Legendre Memory Units) [26], which offer prediction capabilities at least similar to or even better than LSTMs, but also allow implementation in SNNs and therefore deployment on neuromorphic hardware.

Another interesting option for future research is to investigate the performance of our prediction models on other publicly available data sets with particular focus on the composition of the data set. Apart from the imbalanced *NGSIM* data set, possible candidates for this investigation could be the recently proposed *INTERACTION* data set [21] or the trajectory section of Baidu's *ApolloScapes* data set [18]. On the other hand, one could investigate situations other than lane changes such as sudden acceleration or braking, which could also be interesting for learning models to predict, or to further balance the training data of such models. Finally, both data sets used in this work contain mainly highway driving samples.

When moving to more advanced data sets containing urban and interurban driving with a far larger variety of possible scenarios and situations, the question of how to properly balance a trajectory prediction data set for automated driving becomes even more interesting but also more challenging.

REFERENCES

- [1] D. C. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks*, vol. 32, pp. 333–338, 2012.
- [2] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," *arXiv, Computing Research Repository (CoRR)*, vol. abs/1604.07316, 2016. [Online]. Available: <http://arxiv.org/abs/1604.07316>
- [3] A. Lawitzky, D. Althoff, C. F. Passenberg, G. Tanzmeister, D. Wollherr, and M. Buss, "Interactive scene prediction for automotive applications," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, 2013, pp. 1028–1033.
- [4] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH Journal*, vol. 1, no. 1, p. 1, 2014.
- [5] A. Polychronopoulos, M. Tsogas, A. Amditis, and L. Andreone, "Sensor fusion for predicting vehicles' path for collision avoidance systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 549–562, 2007.
- [6] J. Schmüdderich, S. Rebhan, T. Weisswange, M. Kleinhagenbrock, R. Kastner, M. Nishigaki, S. Kusuvara, H. Kamiya, N. Mori, and S. Ishida, "A novel approach to driver behavior prediction using scene context and physical evidence for intelligent adaptive cruise control (i-acc)," in *Future Active Safety Technology Towards zero traffic accidents (FAST-zero)*. FISITA, 2015.
- [7] M. Bahram, C. Hubmann, A. Lawitzky, M. Aeberhard, and D. Wollherr, "A Combined Model- and Learning-Based Framework for Interaction-Aware Maneuver Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1538–1550, 2016.
- [8] S. Bonnin, F. Kummert, and J. Schmüdderich, "A generic concept of a system for predicting driving behaviors," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, 2012, pp. 1803–1808.
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, nov 1997.
- [10] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 961–971.
- [11] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018*, pp. 1468–1476, 2018. [Online]. Available: <http://arxiv.org/abs/1805.06771>
- [12] F. Althe and A. de La Fortelle, "An LSTM network for highway trajectory prediction," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Oct 2017, pp. 353–359. [Online]. Available: <http://dx.doi.org/10.1109/itsc.2017.8317913>
- [13] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1179–1184.
- [14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.
- [15] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollr, "Microsoft COCO: Common objects in context," *arXiv e-prints*, p. arXiv:1405.0312, 2014.
- [16] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [17] 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," *arXiv, Computing Research Repository (CoRR)*, vol. abs/1604.01685, 2016. [Online]. Available: <http://arxiv.org/abs/1604.01685>
- [18] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The ApolloScape dataset for autonomous driving," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018, pp. 954–960.
- [19] F. Mirus, P. Blouw, T. C. Stewart, and J. Conradt, "An investigation of vehicle behavior prediction using a vector power representation to encode spatial positions of multiple objects and neural networks," *Frontiers in Neurobotics*, vol. 13, p. 84, oct 2019.
- [20] J. Colyar and J. Halkias, "US Highway 101 Dataset," 2017. [Online]. Available: <https://www.fhwa.dot.gov/publications/research/operations/07030/index.cfm>
- [21] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clause, M. Naumann, J. Kummerle, H. Königshof, C. Stiller, A. de La Fortelle, and M. Tomizuka, "INTERACTION Dataset: An INTERNATIONAL, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps," *arXiv e-prints*, p. arXiv:1910.03088, Sep 2019. [Online]. Available: <http://arxiv.org/abs/1910.03088>
- [22] C. Eliasmith, *How to build a brain: A neural architecture for biological cognition*. Oxford University Press, 2013.
- [23] R. Gayler, "Vector Symbolic Architectures answer Jackendoff's challenges for cognitive neuroscience," in *ICCS/ASCS International Conference on Cognitive Science*, P. Slezak, Ed., University of New South Wales. CogPrints, 2003, pp. 133–138.
- [24] T. Plate, "Distributed Representations and Nested Compositional Structure," Ph.D. dissertation, University of Toronto, 1994.
- [25] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv, Computing Research Repository (CoRR)*, vol. abs/1803.01271, 2018. [Online]. Available: <http://arxiv.org/abs/1803.01271>
- [26] A. Voelker, I. Kajić, and C. Eliasmith, "Legendre Memory Units: Continuous-Time Representation in Recurrent Neural Networks," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., pp. 15 544–15 553. [Online]. Available: <http://papers.nips.cc/paper/9689-legendre-memory-units-continuous-time-representation-in-recurrent-neural-networks.pdf>