

Neural Motion Prediction for In-flight Uneven Object Catching

Hongxiang Yu, Dashun Guo, Huan Yin, Anzhe Chen, Kechun Xu, Yue Wang and Rong Xiong

Abstract— In-flight objects capture is extremely challenging. The robot is required to complete trajectory prediction, interception position calculation and motion planning in sequence within tens of milliseconds. As in-flight uneven objects are affected by various kinds of forces, motion prediction is difficult for a time-varying acceleration. In order to compensate the system’s non-linearity, we introduce the Neural Acceleration Estimator (NAE) that estimates the varying acceleration by observing a small fragment of previous deflected trajectory. Moreover, end-to-end training with Differentiable Filter (NAE-DF) gives a supervision for measurement uncertainty and further improves the prediction accuracy. Experimental results show that motion prediction with NAE and NAE-DF is superior to other methods and has a good generalization performance on unseen objects. We test our methods on a robot, performing velocity control in real world and respectively achieve 83.3% and 86.7% success rate on a ploy urethane banana and a gourd. We also release an object in-flight dataset containing 1,500 trajectorys for uneven objects.

I. INTRODUCTION

In-flight objects capture is extremely challenging especially for uneven objects. In tens of milliseconds, the robot needs to successively predict the flight trajectory, calculate a feasible interception position and plan the arm’s motion[1, 2, 3]. Among these tasks, the motion prediction is quite critical for the whole catching system. With the predicted trajectory via motion prediction, the arm can be set to the interception position in a suitable catching configuration before the arrival of objects. Otherwise, imprecise prediction with accumulated errors will lead to a imperfect interception position, thus resulting in low success rate of objects capture.

Most of robotic systems are able to catch an even object such as a ball [4, 5, 6, 7, 8], while few studies focus on motion prediction of uneven objects with complex aerodynamics. In these existing studies, they either regard the trajectory of the ball as a parabola [9, 10], or simplify the ball as a solid particle subjects to aerodynamic force and gravity [2, 11, 12]. For example, by formulating the differential equation of the ball’s motion, some research works [13] use the vision system to track the solid ball and then build Extended Kalman filter for state correction. However, we can not transfer the modelling of the ball to uneven object directly, since the dynamics of uneven object is more complicated.

Current modeling for uneven objects can be divided into two categories, physics-based mechanism modeling and traditional machine learning methods. Physics-based modeling[14, 15] requires prior information such as mass, position

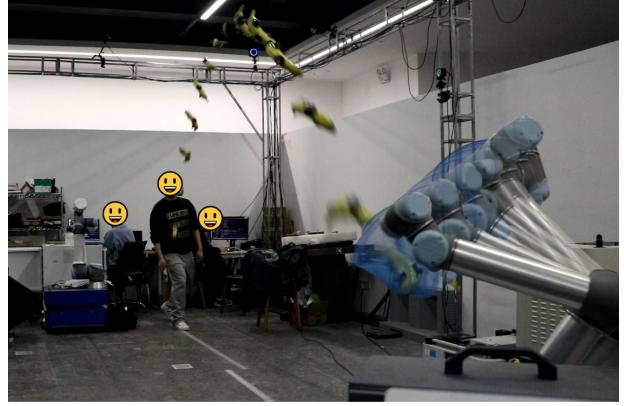


Fig. 1: An uneven banana is thrown towards an UR5 robot 5 meters away at a velocity of 5-6m/s, and the banana only flies for 1s before fallen into UR5’s workspace. By predicting banana’s flying trajectory with NAE-DF, the UR5 catches the banana successfully.

of COM, moment of inertia, and shape of the object, etc. For uneven objects, which have irregular shapes and intricacy aerodynamics, it is troublesome to collect such prior information. When dealing with new objects, these prior information needs to be re-measured or even re-modeled again. As for learning based methods, several works [1, 16, 3] propose to use bundle of machine learning methods to estimate non-linear dynamics after observing some samples of trajectories, which achieve better performance compared to the physics-based modeling. However, the number of learning parameters in kernel methods is highly related to the volume of training data, it cannot be trained with large scale data, thus leading to the limited performance when generalized to unseen objects.

According to the above analysis, we consider that traditional machine learning methods should implicitly capture information associated with external force to show good performance. But they need large model, as they have to capture both temporal and spatial variations, resulting in high complexity. Therefore, we borrow part of the physical idea to formulate the object as a dynamic system for temporal modeling. Then we apply learning method to explicitly estimate acceleration at only one timestep as the system spatial measurement, which is a manifestation of force. Specifically, we propose a neural estimator utilizing a small fragment of previous deflected trajectory to predict the acceleration, named **Neural Acceleration Estimator (NAE)**. Finally, we propose a Differentiable Kalman Filter to filter the state by optimally fusing NAE measurements and temporal dynamics, named **NAE-DF**. Thanks to the differentiability, the whole

¹All authors are with the State Key Laboratory of Industrial Control and Technology, and the Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou 310058, China.

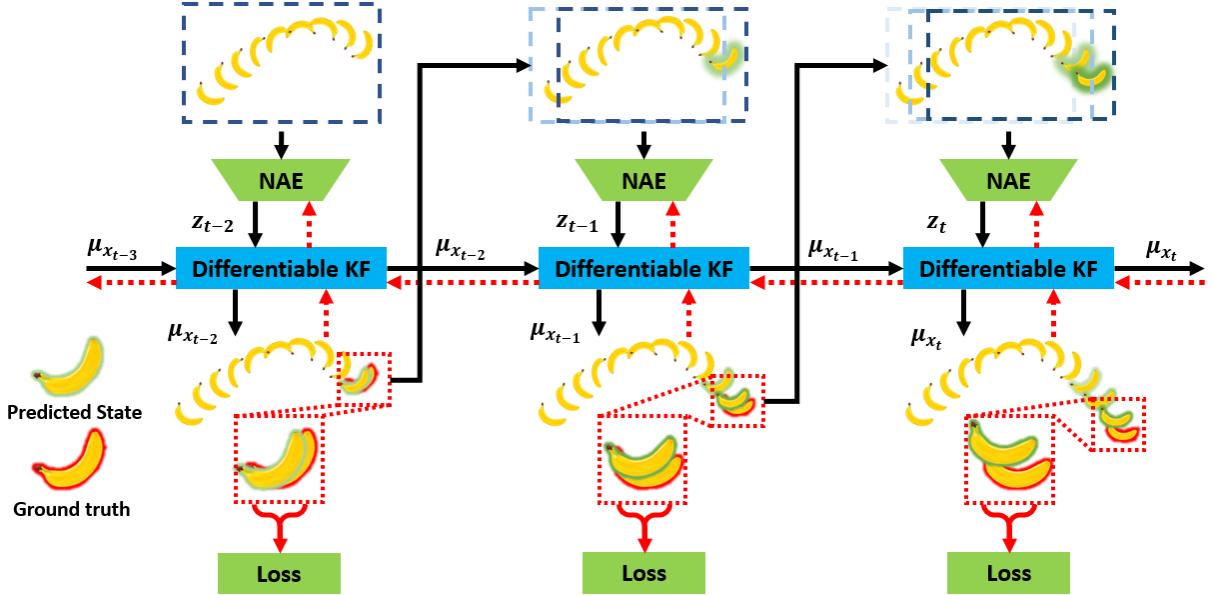


Fig. 2: A integral illustration of NAE-DF. NAE-DF is a Differentiable Filter embedded with a Neural Acceleration Estimator as the measurement model. The entire framework is trained end-to-end, where the black lines indicating the flowing of data and red dash line indicating the backpropagation of loss.

architecture can be trained in an end-to-end manner[17, 18, 19]. In this way, the uncertainty of the NAE measurement can also be indirectly supervised. Such modeling injects the inductive bias to the learning architecture, thereby can be expected to improve the generalization performance.

Experimental results show that the proposed NAE and NAE-DF are superior to existing methods in prediction accuracy and generalization performance for uneven object in both public dataset and real world. Fig.1 shows a demonstration of our catching experiment in real world, and we achieve a success rate of 83.3% for a ploy urethane banana and 86.7% when generalizing the trained banana’s model to an unseen gourd.

Overall, the contributions of this paper are three-fold:

- We propose a Neural Acceleration Estimator which makes accurate estimation to the time-varying acceleration caused by various kinds of external forces, without any prior information like mass, shape or inertia. Then we can estimate the trajectories of in-flight uneven objects with a linear model.
- We embed NAE into a Differentiable Filter and train the NAE-DF in an end-to-end manner, thus supervising the uncertainty of measurement models. Compared to NAE, NAE-DF achieves a better performance for motion prediction on uneven objects.
- The third contribution is the real world validation of NAE and NAE-DF. We use a UR5 manipulator with simple velocity control to perform experiments, and achieve high success rates. Moreover, we open-source a dataset containing more than 1,500 flight trajectories, including the position and the posture of 6 typical objects.

II. RELATED WORK

Catching in-flight objects. Accurately trajectory predicting of the flying object is a significant part in the task of catching in-flight objects. Among the in-flight objects, ball is selected as the experimental object in most tasks since its trajectory is considered to be relatively easy to predict. [2, 12, 10] ignore the effect of air drag and other force applying on the ball and approximate its flight trajectory as a parabola. [5] measures the air drag coefficient in the experimental environment and models the ball accurately in dynamics. [4, 6, 11] use the EKF [13] to correct ball’s flight state. Based on these trajectory prediction methods, the above works have achieved good experimental results catching the ball. But problems come out when these methods are applied to other flying objects.

For the prediction of the trajectory of other flying objects, [14, 15] accomplishes this task by dynamical modeling of the flying objects as well as EKF, but the dynamical modeling requires prior information such as mass, position of COM, moment of inertia, and shape of the objects, which need to be modeled separately for each objects. And it is very difficult to collect these information for uneven objects. [1, 16, 3] estimate the dynamics model of the object by offline learning through the traditional machine learning method, which achieve decent results in predicting the translation and rotation of the object. However, the model parameters of the traditional learning method will augment as the sample size increases, and in addition, the models obtained by traditional learning have poor performance when generalizing to unseen objects.

Differentiable estimator. Conventional Bayes filters have been widely applied for state estimation in robotic applica-

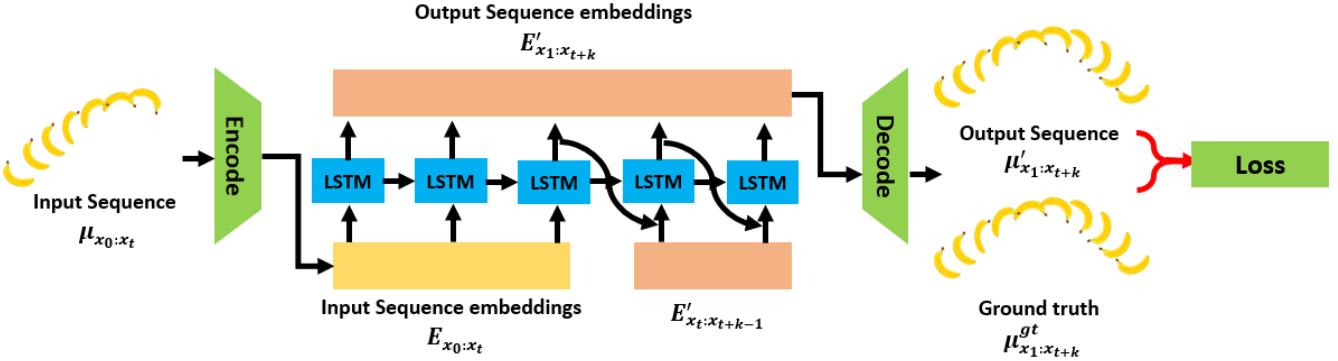


Fig. 3: NAE, an LSTM based acceleration estimator that fusing information from multiple frames. During training, NAE inputs a object state sequence $\mu_{x_0:x_{t-1}}$ from time 0 to t and outputs prediction $\mu'_{x_1:x_{t+k}}$ from time 1 to $t-1+k$. Three loss functions are defined to train the NAE model, respectively for learning single-step prediction, multi-step prediction and embedding reconstruction. During using, NAE inputs a object state sequence $\mu_{x_0:x_t}$ from time 0 to $t-1$ to estimate the object state μ_{x_t} at time t .

tions [20]. On the other side of front-end, the latest deep learning technique brings more precise measurement with sensors [21]. In this context, the combination of the state estimation and deep learning is becoming a research focus in recent years.

To achieve this, Haarnoja et al. [17] proposed a backprop Kalman filter system, in which the state estimator and recurrent network are trained together with backpropagation. The authors demonstrated the effectiveness with synthetic tracking task and visual odometry in the real world. As for the range sensors, the differentiable Kalman filter can also be integrated to the end-to-end system for vehicle pose tracking [18]. Despite the Kalman filter above, some research works focused on building end-to-end particle filter [22, 23]. However, since the resampling step is non-differentiable, most of particle filter based methods only used the network as measurement model [24], and the whole system is not trained in end-to-end manner. In [23], a soft-sampling policy is introduced to address the issue, thus making the differentiable particle filter feasible for state estimation.

III. METHODS

The flight of uneven object is a dynamic system constructed of position, velocity, and acceleration. In some works, all other external forces are not considered. The acceleration is fixed to only gravity, and the order of the system can be reduced, leading to the prediction error. For a better physical model, the acceleration is regarded time-varying, but it lacks real-time measurement, thus is modeled for a specific object in prior, failing to be generalized to new object. Following this idea, our model uses NAE to online measure the acceleration, and update the state, achieving accuracy and generalization at the same time.

In order to deal with the system's non-linearity caused by air drag, air lift and Magnus force, etc.[15], Neural Acceleration Estimator(NAE), as shown in Fig. 3, explicitly estimates time-varying acceleration by a Long Short-Term Memory network. Taking both prediction result given by

linear propagation model and the measurement made by NAE into consideration, we supervise the uncertainty of measurement step by end to end training NAE with a Differentiable Kalman Filter. Fig. 2 shows the integral diagram of NAE-DF. We introduce NAE in Section III-A and give more detail about NAE-DF model in Section III-B, as shown in Fig. 4

A. Neural Acceleration Estimators

Uneven objects are affected by various kinds of forces, which leads to a non-linear dynamic system. This non-linearity changes acceleration all the time during object's flight. In-flight acceleration is hard to be measured by sensors but can be estimated by observing a small fragment of previous deflected trajectory, which is accomplished by the Neural Acceleration Estimators.

As a non-linear model, the LSTM network is qualified for fusing information from multiple frames and learning long-term dependence. As shown in Fig. 3, we define the object state $\mu_{x_t} \in \mathbb{R}^9$ at frame t as a nine-dimensional vector composed of the current position, velocity and acceleration. The NAE model inputs a object state sequence $\mu_{x_0:x_t}$ from frame 0 to $t-1$ to estimate the object state μ_{x_t} at next frame t . We first map the input state sequence to a high-dimensional space $E_{x_t} \in \mathbb{R}^{128}$ using an Encoder consist of fully connected layers, and use \tanh as the activation function. After processing, the embedded output $E'_{x_{t+1}}$ is remapped back to $\mu'_{x_{t+1}}$ in the original space through a Decoder as the final result of NAE. The core of NAE is an LSTM-based multi-step prediction model. We use three loss functions for training.

One-step Teacher forcing loss: One-step Teacher forcing loss helps NAE learn single-step prediction. Given the sequence of object state $\mu_{x_0:x_t}$ from frame 0 to $t-1$, NAE outputs the sequence $\mu'_{x_1:x_{t+1}}$ from frame 1 to t . One-step Teacher forcing minimizes the MSE loss between the prediction sequence $\mu'_{x_1:x_{t+1}}$ and the ground truth sequence $\mu^{gt}_{x_1:x_{t+1}}$,

$$\mathcal{L}_1 = \frac{1}{t} \sum_{i=0}^{t-1} \left\| \mu_{x_{i+1}}^{gt} - \mu'_{x_{i+1}} \right\|^2 \quad (1)$$

Multi-step Free running loss: Multi-step Free running helps NAE learn multi-step prediction. Given the sequence of object state $\mu_{x_0:x_t}$ from frame 0 to $t-1$, NAE outputs the object state sequence $\mu'_{x_1:x_{t+1+k}}$ from frame 1 to $t+k$ after k prediction steps. For the frame with a given state, the input of the LSTM unit is the mapping E_{x_t} of the known state μ_{x_t} at the corresponding frame, and for the frame without given state, the input is the output $E'_{x_{t-1}}$ of the LSTM unit at the previous frame $t-1$. Multi-step Free running also uses MSE loss to minimize the error between the prediction sequence $\mu'_{x_{t+1}:x_{t+1+k}}$ and the ground truth sequence $\mu_{x_{t+1}:x_{t+1+k}}^{gt}$,

$$\mathcal{L}_2 = \frac{1}{k} \sum_{i=t}^{t+k} \left\| \mu_{x_{i+1}}^{gt} - \mu'_{x_{i+1}} \right\|^2 \quad (2)$$

Reconstruction loss: Reconstruction step learns the Encoder and Decoder composed of a fully connected layers. The Encoder maps the input sequence $\mu_{x_0:x_t}$ from 0 to $t-1$ to the high-dimensional space, then the Decoder maps it back to the original space, minimizing the reconstruction loss,

$$\mathcal{L}_3 = \frac{1}{t} \sum_{i=0}^{t-1} \left\| \mu_{x_i}^{gt} - D(E(\mu_{x_i})) \right\|^2 \quad (3)$$

B. Neural Acceleration Estimator with Differentiable Filter

Uncertainty helps us to make an optimal fusion of prediction and measurement by estimating a joint probability distribution. Since uncertainty of both prediction and measurement are lack of supervision, we embed NAE into a Differentiable Filter, which takes the acceleration estimated by NAE as the measurement as shown in Fig. 2. In the learning stage, we train the NAE-DF in an end-to-end manner and a maximum likelihood formulation.

Kalman Filter models statistical noise and other inaccuracies and is a general technique to fuse the sensor data in sequential[18]. In prediction step, known dynamical model will propagate the current state to the next state, during which the uncertainty increases. Then the measurement step makes estimation more confident by multiplying two Gaussian distributions. Fig. 4 gives the internal details of the DF module. As we put all complex non-linearity caused by air drag, air lift, Magnus force and gravity, etc. into the NAE's measurement, we can use a simple linear propagation model derived from gravity as the prediction model. Therefore the prediction step is formulated as follows:

$$\bar{\mu}_{x_t} = A\mu_{x_{t-1}} \quad (4)$$

$$\bar{\Sigma}_{x_t} = A\Sigma_{x_{t-1}}A^T + Q \quad (5)$$

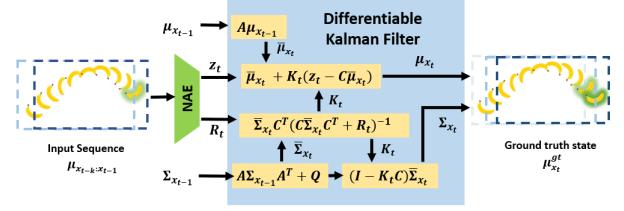


Fig. 4: The internal details of the Differentiable Filter. The model consists of a prediction step with linear dynamical model and a measurement step with non-linear NAE.

where the matrix A is

$$A = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 & \frac{1}{2}\Delta t^2 & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 & 0 & \frac{1}{2}\Delta t^2 & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t & 0 & 0 & \frac{1}{2}\Delta t^2 \\ 0 & 0 & 0 & 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and Q is a pre-defined noise covariance matrix.

For the measurement step, we use NAE as the observation model whose input is a recent sample of object in-flight trajectory and output is an observation of the current state of the object. The covariance matrix R_t of the current state is also supervised by end-to-end training. Measurement step is formulated as follows:

$$K_t = \bar{\Sigma}_{x_t} C^T (C\bar{\Sigma}_{x_t} C^T + R_t)^{-1} \quad (6)$$

$$\mu_{x_t} = \bar{\mu}_{x_t} + K_t (z_t - C\bar{\mu}_{x_t}) \quad (7)$$

$$\Sigma_{x_t} = (I - K_t C) \bar{\Sigma}_{x_t} \quad (8)$$

where C is the observation matrix. Since we only take the acceleration term of the NAE's estimated state as the observation, the matrix C is defined as:

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

From a probability point of view, Differentiable Filter maximizes the data likelihood $\mathcal{N}(\mu_{x_t}, \Sigma_{x_t})$, so we can train NAE-DF by maximizing the posterior probability, as shown in Fig. 2:

$$\text{maximize } \mathcal{N}(\mu_{x_t}^{gt}; \mu_{x_t}, \Sigma_{x_t}) \quad (9)$$

We apply negative log-likelihood to formulate the loss function:

$$\mathcal{L}_4 = \frac{1}{k} \sum_{t=k}^t (\mu_{x_t}^{gt} - \mu_{x_t})^T \Sigma_{x_t}^{-1} (\mu_{x_t}^{gt} - \mu_{x_t}) + \gamma \det(\Sigma_{x_t}) \quad (10)$$

where γ is a constant factor for regulation.

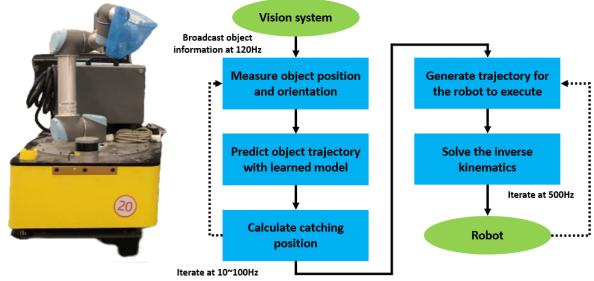


Fig. 5: The system flow and hardware equipment we use to form our real world catching system. A motion capture system provides vision information including object's position and orientation at 120 Hz. The interception position is calculated by learned model, then the arm following a velocity control law will move to the target position for capture.

IV. SYSTEM OVERVIEW

Fig. 5 illustrates the system flow and hardware equipment of the real world catching system we built to verify our algorithm. First, we collect the flight information of uneven objects through a motion capture system. Next, two independent threads work simultaneously to generate the interception position and control the UR5 respectively. The interception position generation thread predicts the trajectory of the flying object through the model, and generates the appropriate interception position by calculating the intersection of the trajectory and the arm's workspace. According to the generated position, the robot arm control thread calculates the motion trajectory by inverse kinematics, and corrects the position through velocity control during reestimation. The velocity control law is defined as follows:

$$v = \frac{1 - e^{-kd}}{1 + e^{-kd}} v_{max} \quad (11)$$

where $d = \|\mathbf{p}_{cur} - \mathbf{p}_{tar}\|^2$ is the distance between the arm's current position \mathbf{p}_{cur} and the target position \mathbf{p}_{tar} , k is a proportional coefficient used to accelerate the convergence of the UR5, and v_{max} is the maximum velocity of UR5 executor. In our experiment we chose $k = 12$ and $v_{max} = 1.85m/s$.

V. EXPERIMENTAL RESULT

In this section, we carried out a series of experiments to evaluate our method. The goals of the experiments are:

- to demonstrate our method is capable of predicting more accurate trajectories of in-flight uneven objects than traditional learning methods.
- to indicate that our method is effective and efficient for the task of catching in-flight uneven object with velocity of 5-6m/s.
- to test the generalization to unseen objects of our method.

We compare our method with two methods: 1) Newton: a method which regards the trajectory of the in-flight uneven



Fig. 6: The objects in the left are uneven objects we used for data collection. Infrared reflectors are fixed on objects for motion capture. Blue spoon net in the right is used as the basket. The Radius of spoon net is 10 cm.

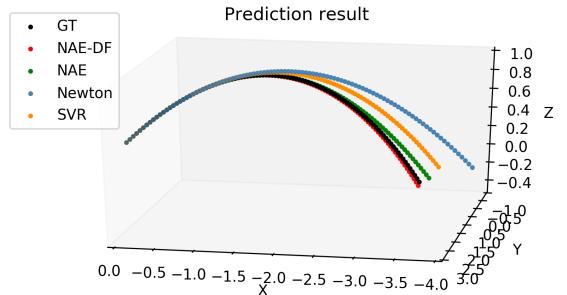


Fig. 7: Prediction result of the four algorithms for a trajectory in banana testing set. NAE-DF's prediction(red) is the most similar to the ground truth(black) trajectory.

objects as a parabola. 2) Support Vector Regression (SVR) [1, 16]: a method using bundle of machine learning methods to estimate non-linear dynamics learning from some samples of trajectories.

A. Dataset Experiment

We collect over 1,500 trajectories of 6 typical objects, 90% of which for training set and 10% for testing set. See Fig. 6 for more details. All vegetable models are made of PU(ploy urethane) materials, including a banana, a bamboo, a green and a bitter gourd. Besides we also have a toy Paige, and a bottle with water. The in-flight dataset are generated with data augmentation methods such as translation and rotation around Z axis so models can adapt to various inputs. We train the SVR, NAE, and NAE-DF models on the training set, and compare their prediction precision on the testing set. In dataset evaluation, the goal of the trajectory is set to the last frame's position.

Prediction Result. Fig. 7 shows the prediction result of the four algorithms on a trajectory in banana's testing set. The diagram shows that the error of direct propagation using gravity acceleration(Newton's method[9, 10]) is very large, followed by the SVR and NAE. NAE-DF's prediction is the most similar to the ground truth trajectory. Fig. 8 and Fig. 9 show the accumulated error decreases when there are

TABLE I: Leading Time Criterion defined by how early achieve the desired precision as 1cm in Cartesian space. Larger value means higher prediction accuracy. NAE-DF gets the highest value, followed by NAE.

	Bottle(half)	Bamboo	Banana	Green	Gourd	Paige
Newton	0.15±0.03	0.11±0.02	0.10±0.02	0.08±0.01	0.10±0.02	0.09±0.01
SVR	0.12±0.08	0.14±0.15	0.16±0.09	0.04±0.08	0.09±0.11	0.14±0.11
NAE (Ours)	0.21±0.08	0.23±0.05	0.25±0.08	0.17±0.06	0.25±0.06	0.26±0.07
NAE-DF(Ours)	0.25±0.09	0.30±0.08	0.30±0.09	0.20±0.04	0.26±0.07	0.28±0.10

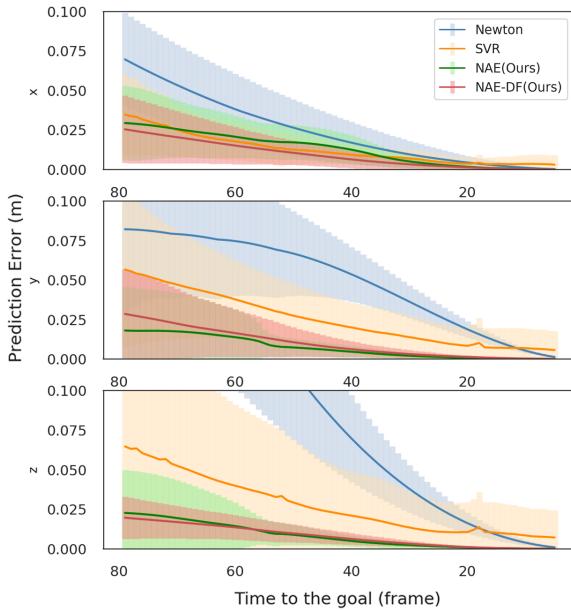


Fig. 8: The accumulated error decreases when there are fewer frames to predict. Horizontal coordinate represents remained frames to the goal. Vertical coordinate represents the L_2 norm with goal position's coordinates. In the early time, NAE and NAE-DF get smaller accumulated error and standard deviation than the other two methods.

fewer frames to predict. In the early time, NAE and NAE-DF get smaller accumulated error and standard deviation than the other two methods. This suggests that our time-varying acceleration estimation is beneficial to improve the trajectory prediction accuracy. Besides, NAE-DF can achieve a better performance than NAE because the end-to-end training with Differentiable Filter can provide stronger constraints.

Leading Time Evaluation. Kim et al.[16] define an evaluation criterion for different algorithms that is how early achieve the desired precision as 1cm in Cartesian space, called Leading Time. This is also the desiderate to compensate lag of execution time for a robot. The larger the value is, the better the model performs. Table. I indicates that for a banana, NAE-DF can obtain an estimation results less than 1cm error with 0.30s ahead, and the value for NAE, SVR, and Newton' method are 0.25s, 0.16s, and 0.10s respectively. This is because the flight of the uneven object is a highly non-linear system. In contrast, NAE and NAE-

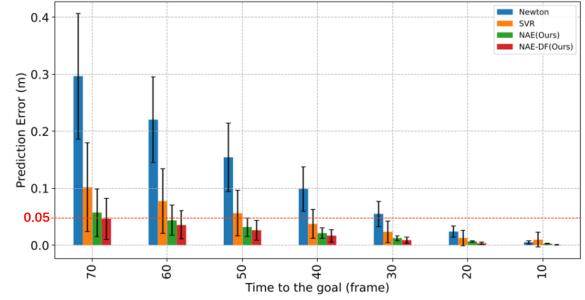


Fig. 9: The accumulated error decreases when there are fewer frames to predict. Horizontal coordinate represents remained frames to the goal. Vertical coordinate represents the L_2 norm with ground truth of goal position. In the early time, NAE and NAE-DF get smaller accumulated error and standard deviation than the other two methods. NAE-DF's error stays under 0.05m.

DF are capable of modeling this non-linearity thus performs more efficiently in this experiment.

Leading Time Evaluation for Generalization. In order to verify the generalization performance, we use the Leading Time criterion mentioned above for comparison. As shown in Fig. 10, the proposed NAE and NAE-DF are superior to Newton's method and SVR in generalization performance for unseen objects. The generalization performance of NAE-DF on objects other than the bottle is better than NAE. This is due to the inductive bias injected to the learning architecture by NAE and NAE-DF.

B. Real World Experiment

Fig. 11 gives an illustration of our real world experiments. We try to capture a PU banana, a PU bitter gourd and a bottle with water by the real world system. Except those throws that don't intersect with the arm's workspace, the NAE-DF model trained on banana's data has a success rate of 83.3% when catching a flying banana, while 73.3% for NAE, 56.7% for SVR, 40.0% for Newton method, which validates effectiveness of our method for catching in-flight uneven objects in real world experiments. Then we test the generalization performance of the model trained with banana data by catching the unseen bitter gourd and a bottle with water. The success rates of each generalization experiments are 86.7% and 53.3% respectively, indicating that our model has a decent generalization performance.

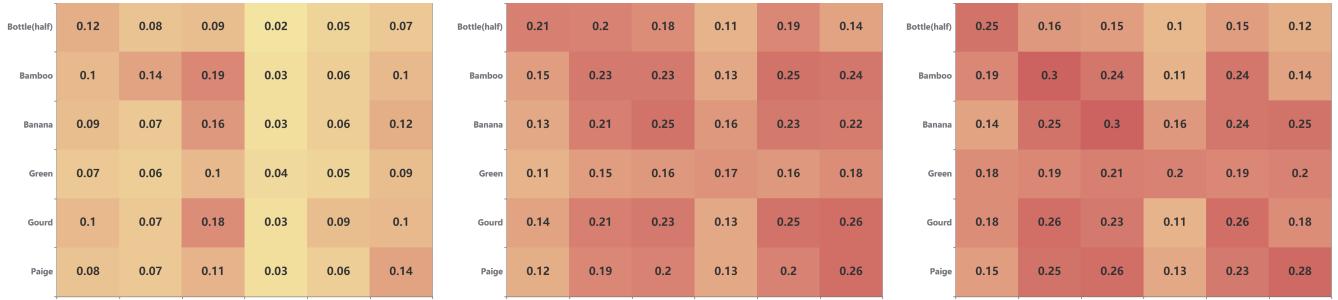


Fig. 10: Generalization performance indicated by Leading Time. x coordinate represents the dataset we test. y coordinate represents the dataset we used to train the model.

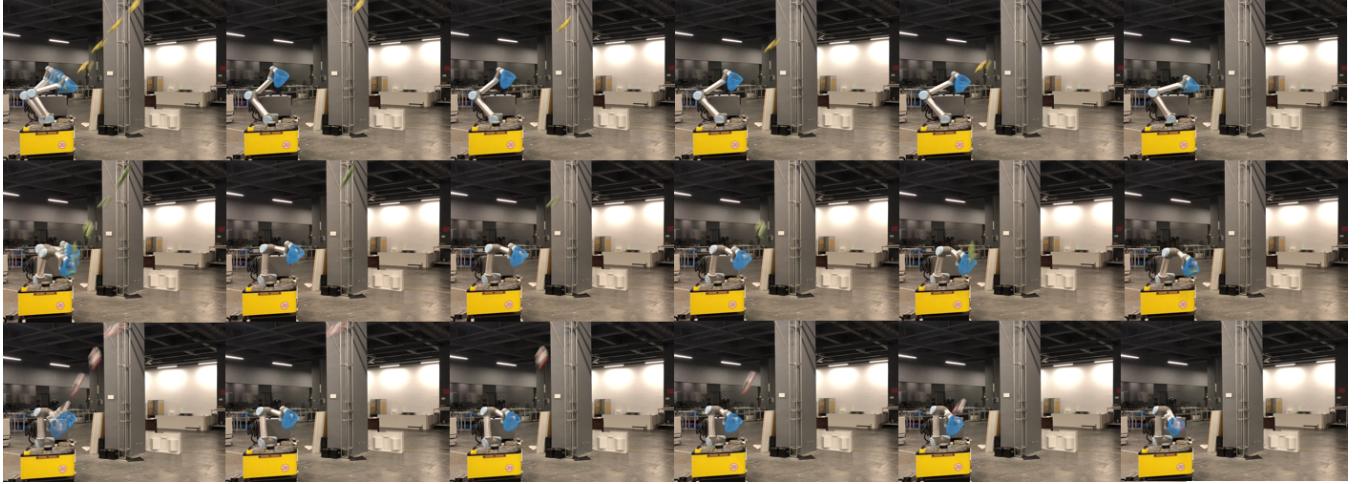


Fig. 11: Success cases for capture a PU banana, a PU bitter gourd and a bottle with water. The used model is NAE-DF trained on banana's data.

C. Failure Case Analysis

The typical failure cases in the real world experiment are shown in the Fig. 12. The main reason that leads to a failure is that currently we don't consider the optimal interception pose, but only the interception position. The radius of the basket is relatively small (10cm, see Fig. 6 for more detail) for some objects' poses. In these cases, objects will be bounced off by the basket. In the following work, we will carry out the motion prediction for in-flight pose, which will optimize the interception pose.

VI. CONCLUSIONS

Tackling the challenging task of motion prediction for in-flight uneven object, we propose a Neural Acceleration Estimator that measures the time-varying acceleration caused by system's non-linearity. Furthermore, we embed NAE into a Differentiable Filter which is trained in the end-to-end manner to give a supervision to uncertainty in measurement. We verify the effectiveness of the algorithm on the dataset as well as on the real world system. With simple velocity control, it is possible to achieve interception of a variety of objects. Compared with the existing uneven objects' motion

prediction methods, our algorithms have great advantages in prediction accuracy and generalization performance. At the same time, we open-source an in-flight object dataset. In the future work, we will predict the flying pose of the objects and optimize the motion planning of the robotic arm for further improvement.

REFERENCES

- [1] Seungsu Kim, Ashwini Shukla, and Aude Billard. "Catching objects in flight". In: *IEEE Transactions on Robotics* 30.5 (2014), pp. 1049–1065.
- [2] Jie Chen, Shen Shen, and Henry YK Lau. "Hitting flying objects with learning from demonstration". In: *2017 18th International Conference on Advanced Robotics (ICAR)*. IEEE. 2017, pp. 55–60.
- [3] Seyed Sina Mirrazavi Salehian, Mahdi Khoramshahi, and Aude Billard. "A dynamical system approach for softly catching a flying object: Theory and experiment". In: *IEEE Transactions on Robotics* 32.2 (2016), pp. 462–471.



Fig. 12: Typical failure case. Objects are bounced off by the basket for a sub-optimal interception pose.

- [4] Ke Dong et al. “Catch the Ball: Accurate High-Speed Motions for Mobile Manipulators via Inverse Dynamics Learning”. In: *arXiv preprint arXiv:2003.07489* (2020).
- [5] Berthold Bäuml et al. “Catching flying balls with a mobile humanoid: System overview and design considerations”. In: *2011 11th IEEE-RAS International Conference on Humanoid Robots*. IEEE. 2011, pp. 513–520.
- [6] Berthold Bäuml, Thomas Wimböck, and Gerd Hirzinger. “Kinematically optimal catching a flying ball with a hand-arm-system”. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2010, pp. 2592–2599.
- [7] Roberto Lampariello et al. “Trajectory planning for optimal robot catching in real-time”. In: *2011 IEEE International Conference on Robotics and Automation*. IEEE. 2011, pp. 3719–3726.
- [8] Yifeng Zhang et al. “Spin observation and trajectory prediction of a ping-pong ball”. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2014, pp. 4108–4114.
- [9] Won Hong and Jean-Jacques E Slotine. “Experiments in hand-eye coordination using active vision”. In: *Experimental Robotics IV*. Springer, 1997, pp. 130–139.
- [10] Marcia Riley and Christopher G Atkeson. “Robot catching: Towards engaging human-humanoid interaction”. In: *Autonomous Robots* 12.1 (2002), pp. 119–128.
- [11] Udo Frese et al. “Off-the-shelf vision for a robotic ball catcher”. In: *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*. Vol. 3. IEEE. 2001, pp. 1623–1629.
- [12] Mark Müller, Sergei Lupashin, and Raffaello D’Andrea. “Quadrocopter ball juggling”. In: *2011 IEEE/RSJ international conference on Intelligent Robots and Systems*. IEEE. 2011, pp. 5113–5120.
- [13] Allen L Barker, Donald E Brown, and Worthy N Martin. “Bayesian estimation and the Kalman filter”. In: *Computers & Mathematics with Applications* 30.10 (1995), pp. 55–77.
- [14] Yan-Bin Jia, Matthew Gardner, and Xiaoqian Mu. “Batting an in-flight object to the target”. In: *The International Journal of Robotics Research* 38.4 (2019), pp. 451–485.
- [15] Matthew Gardner and Yan-Bin Jia. “Motion Estimation of Free-Flying Objects: Aerodynamics, Constrained Filtering, and Graph-Based Feature Tracking”. In: () .
- [16] Seungsu Kim and Aude Billard. “Estimating the non-linear dynamics of free-flying objects”. In: *Robotics and Autonomous Systems* 60.9 (2012), pp. 1108–1122.
- [17] Tuomas Haarnoja et al. “Backprop KF: learning discriminative deterministic state estimators”. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. 2016, pp. 4383–4391.
- [18] Huan Yin et al. “RaLL: End-to-end Radar Localization on Lidar Map Using Differentiable Measurement Model”. In: *IEEE Transactions on Intelligent Transportation Systems* (2020).
- [19] Michelle A Lee et al. “Multimodal Sensor Fusion with Differentiable Filters”. In: *arXiv preprint arXiv:2010.13021* (2020).
- [20] Sebastian Thrun. “Probabilistic robotics”. In: *Communications of the ACM* 45.3 (2002), pp. 52–57.
- [21] Changhao Chen et al. “A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence”. In: *arXiv preprint arXiv:2006.12567* (2020).
- [22] Rico Jonschkowski, Divyam Rastogi, and Oliver Brock. “Differentiable particle filters: End-to-end learning with algorithmic priors”. In: *arXiv preprint arXiv:1805.11122* (2018).
- [23] Peter Karkus, David Hsu, and Wee Sun Lee. “Particle filter networks with application to visual localization”. In: *Conference on Robot Learning*. PMLR. 2018, pp. 169–178.
- [24] Huan Yin et al. “Locnet: Global localization in 3d point clouds for mobile vehicles”. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2018, pp. 728–733.