IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA MINING FOR INTERNET OF THINGS

Received September 10, 2019, accepted October 4, 2019, date of publication October 8, 2019, date of current version October 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946186

A Reinforcement One-Shot Active Learning Approach for Aircraft Type Recognition

HONGLAN HUANG[1], YANGHE FENG[1], JINCAI HUANG[1], JIARUI ZHANG[2], AND LI CHEN[1]

[1]College of Systems Engineering, National University of Defense Technology, Changsha 410073, China
[2]College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China

Corresponding author: Yanghe Feng (fengyanghe@nudt.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 71701205 and Grant 71701206.

ABSTRACT Target recognition is an important aspect of air traffic management, but the study on automatic aircraft identification is still in the exploratory stage. Rapid aircraft processing and accurate aircraft type recognition remain challenging tasks due to the high-speed movement of the aircraft against complex backgrounds. Active learning, as a promising research topic of machine learning in recent decades, can use less labeled data to obtain the same model accuracy as supervised learning, which greatly reduces the cost of labeling a dataset. Instead of manually developing policies of accessing the labels of desired instances, an improved active learning approach, which can not only learn to classify samples using small supervision but additionally capture a relatively optimal label query strategy, was developed by employing the reinforcement learning in the process of decision-making. The proposed model was first tested with the Amsterdam Library of Object Images (ALOI) dataset and then used to perform aircraft type recognition on one-month real-world flight track data. Our method offers a satisfactory solution for learning new concepts rapidly from a small amount of data, which well meets the needs of aircraft type recognition task in practical application.

INDEX TERMS Aircraft type recognition, active learning, cross entropy, one-shot learning, reinforcement learning.

I. INTRODUCTION

With the rapid increase in the variety and quantity of aircraft, precise identification of aircraft types is not only an important task of air traffic control in daily life but also a vital military mission. However, aircraft type recognition methods are still in the exploratory stage, and mature aircraft recognition theories and systems have not yet been formed. In order to achieve better recognition accuracy, aircraft type recognition work still requires substantial human input. As a hot topic in both academia and industry, machine learning has made major advances in areas such as pattern analysis [1], image processing and natural language processing. Therefore, the use of machine learning methods to reduce the workload of human experts in aircraft type recognition tasks has become a meaningful research direction.

For many real-world tasks, labeled data are scarce whereas unlabeled data are abundant [2]. As is widely acknowledged in this domain, formulating labels is a straightforward strategy to process data that involves plenty of human interaction. It is relatively easy to obtain a large number of unlabeled instances while acquiring labeled instances is expensive (e.g. manually annotated) and is not always available in large volumes [3], [4]. Prior investigations have demonstrated that accessing the ground-truth label of a dataset not only requires the effort of considerable experts in related fields but also takes more than 10 times longer to label a sample than to collect it [5]. As dataset volumes grow continuously, the learning systems tend to generalize better, but the cost of annotation has also increased dramatically [6]. To achieve better recognition accuracy, aircraft type recognition work still requires considerable participation of human experts, since labeling is typically done manually, considered to be time-consuming and labor-intensive. Thus, there is a strong demand for training an accurate machine learning model to mitigate the heavy workload of human experts in aircraft type recognition tasks.

As a promising approach to this goal, active learning is a widely applicable machine learning framework that serves to reduce the cost of annotation without sacrificing model

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Wei Tsai.

performance [7]. Human learning process is simulated by active learning approaches in some way: it iteratively queries the labels of certain instances and adds them to the training set, and tries to improve the generalization performance of the model with fewer queries. This method has been well studied during the past years and benefited a variety of practical scenarios, like information retrieval [8], image and speech recognition [9], text analysis [10]–[12] and automatic target recognition [13].

Humans are able to learn and generalize new concepts from only a few labeled instances [14], [15]. One-shot (or few-shots) learning simulates this process in the literature to some extent [15]. Inspired by this, we aimed to design an artificial intelligence agent that could inherit similar capabilities and pose fewer requests for the labels of new instances during the training process [16]. In active learning, an ideal situation is that labeling critical instances is still required but the number of queries can be minimized. Thus, we prefer to study a problem at the crossroad of active learning, reinforcement learning and one-shot learning [17], [18] rather than a human-designed criterion. More specifically, the selection or design of the strategy of labeling new instances can be performed automatically.

Our study introduces a novel learning model that not only learns to classify samples using small supervision but additionally captures a relatively optimal label query strategy. We treat active learning method as a meta-learning problem [19] and train this active query strategy network with reinforcement learning. Mostly inspired by the work of Woodward and Finn [20] and Huang *et al.* [21], this paper can be viewed as a practical extension. We study the case of streamed-based setting where the model considers a stream of instances and needs to classify one sample after another. It's a natural fit for an active learner using reinforcement learning to solve a continuous decision problem, since the next decision is affected by the previous action (when and which instance to query next depending on the current state of the basic learner). Therefore, a cogent nonmyopic strategy can be learned by the active query system trained by reinforcement learning, and effective decisions can be made with little supervision.

In particular, our contributions in this paper can be summarized as follows:

1) We address the challenge of aircraft type recognition in practical application and design the aircraft type recognition task in a novel stream-based online learning way. We collect one month's worth of flight track data in a real-world environment, not in a simulated environment, and greater quantities of flights and types of aircraft are considered than previous studies.
2) We employ a novel reinforcement one-shot active learning approach [21] to the task of object recognition using Amsterdam Library of Object Images (ALOI) dataset [22] and the aircraft dataset. It is sought to be the first time considering the issue of how an aircraft recognition system can improve performance under limited resources by this meta-learning approach.
3) Compared to various state-of-the-art algorithms, we experimentally demonstrate the efficiency of present method in exploring label query strategies based on the uncertainty [23] of instances end-to-end and its ability to learn new concepts rapidly from a small amount of data, which well meets the needs of practical applications.

II. RELATED WORK

For now, investigation on aircraft automatic recognition is still in the exploration stage, and most of the existing studies focus on the method of graphic image processing [24]–[28]. Radar signal analysis has also been widely used in air traffic management [29]. Image-based methods and radar-based methods primarily use features of aircraft profiles to identify the type of aircraft. Aircraft recognition based on contour is mainly to find the approximate invariant features [30]–[32]. Commonly used invariant feature extraction methods include Hu matrix [31], affine distance, Fourier descriptor [33], wavelet moment, and Zernike moments. However, contour-based methods may encounter some inherent deficiencies [1]. In real-time applications, one common technique for identifying military aircraft is Identification Friend Foe (IFF). Civil aircraft uses an IFF-like technique called Secondary Surveillance Radar (SSR) [29]. The fundamental disadvantage of technologies such as IFF and SSR is the need for active pilot cooperation, which makes these technologies inefficient and less practical.

Aiming at lowering the cost of annotation without sacrificing model performance, active learning as a subfield of machine learning has been well studied during the past years [9], [34], [35]. The idea of active learning benefits a variety of practical scenarios, including film recommendation [36]–[38], medical image classification [39], natural language processing and so on. A common view of choosing the appropriate instance for labeling is based on maximizing the expected informativeness for labeled instances [40]. Uncertainty sampling [41] is one of the most popular active learning methods, in which the classifier selects the sample with the highest measure of uncertainty to query. Query by committee is another well-motivated active learning framework, in which a committee of classifier is trained on the same data set, and the next query is chosen according to the principle of maximal disagreement [42], [43]. Ebert *et al.* proposed a diversity promoting sampling method that uses graph density to determine most representative points [44]. Konyushkova *et al.* proposed a data-driven approach called Learning Active Learning, and the key idea is to train a regressor that predicts the expected error reduction for a candidate sample in a particular learning state [45]. In general, most of these strategies rely heavily on heuristics or theoretical measures, such as similarity measures between previous and current instances [46], or the extent of uncertainty in label prediction [46]–[48]. However, heuristic-based

active learning methods may fail when the data distribution of the underlying learning problems varies (e.g. a new category appears).

To move away from engineered selection heuristics, we cast active learning as a decision process, and use reinforcement learning to learn an action policy for an active learner. The premise of active learning is that costs associated with label requests and making false predictions can be reasonably modeled [20]. Those costs can be optimized by reinforcement learning through explicitly setting reward and punishment, and an action strategy can be directly determined. Thus, we believe that the combination of reinforcement learning and active learning is a reasonable and appealing approach to stream-based online cases. Some recent studies have also generated interest in a similar idea. Woodward and Finn [20] firstly focused on learning an optimal policy for active learning task with the help of reinforcement learning. They use reinforcement learning with a recurrent-neural-network-based Q function in a sequential one-shot learning task to decide between predicting a label and acquiring the true label at a cost [7]. Bachman *et al.* [2] and Pang *et al.* [19] studied a pool-based active learning algorithm in a meta-learning fashion. Puzanov and Cohen [16] developed an artificial intelligence classification systems using the same idea. Recent methods such as meta-learning and one-shot learning are closely related to our model [15]. A supervised meta-learning model based on memory-augmented neural networks was proposed by Santoro *et al.* [49], which focused on the same learning task as ours.

III. MODEL DESCRIPTION

The framework of our proposed reinforcement one-shot active learning (ROAL) method is presented in this section. We mainly consider a single pass stream-based online active learning scenario, in which the model decides, while observing instances continuously obtained from the data stream and presented in an exogenously-determined order, whether to predict each instance's label or to pay a cost to query its label. The learner usually observes one unlabeled instance from a continuous stream each cycle and has to choose the appropriate action (predict the label or query the label) for each instance of the arrival [40]. A deep recurrent neural network [50] function approximator is used to act as a function approximator for a Q-network, and the output of the network is connected to a fully connected layer, which produces the actual Q-values. Moreover, the cross entropy [51] term is employed in the loss function to improve the performance of the classifier.

A. TASK DESCRIPTION

Obtaining the ground-truth label of a data instance is time-consuming and expensive in the scenario of stream-based online learning. Therefore, judiciously identifying the number of instances to label is in urgent need for the classification algorithm [35], [52]. Under the setting of this [35], [53],

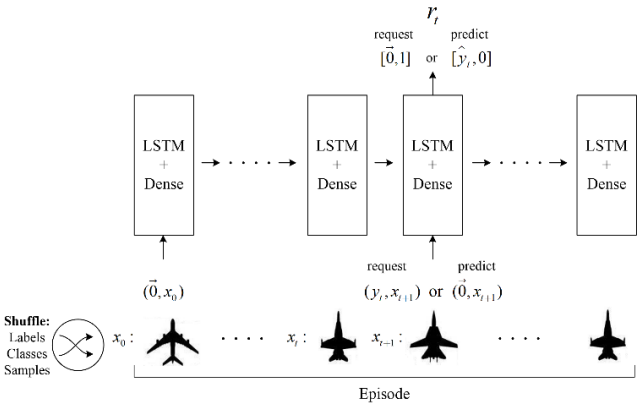


FIGURE 1. Task structure. For instances in the datasets, the classes and their labels, as well as specific samples of each class are shuffled and randomly presented at each episode.

the algorithm makes a decision, whether to request the ground truth label when instance arrives. The classification task that we focus on is a stream of instances (e.g. images or aircraft target track) for which labels must be queried or predicted. In the setting of one-shot learning [15], [49], in order to maximize the performance of the model on the new classes that are not present in the training set, the performance of the model is improved over short training episodes and a small number of instances per class. The structure of the active learning task we propose is shown in Fig. 1. At each time step of the episode, an instance x_t is given to the model, and the model needs to decide an action to take. Assuming that there are up to N possible classes in each episode, the action space is defined as following:

$$\mathcal{A} \triangleq c_1, \ldots, c_N, a_{req} \tag{1}$$

Let a_t be the action that the model takes at time step t. When the model predicts the label of the instances as one of N possible classes (e.g. class i) without requiring the ground truth of the label at time t, action $a_t = c_i$ is taken. When the model requests the true label y of the instance, action $a_t = a_{req}$ is taken. The action a_t is represented by a one-hot vector which the first N bits are consistent with the optionally predicted label $\hat{y}$ and are followed by a bit for requesting the label. The model can only perform one action at a time step, either predict the label of the instance or request the label, since only one bit of the vector can be 1. If the model queries the label of instance x_t, then no other action (prediction) will be made, and the true label y_t will be sent to the model along with a new instance x_{t+1} at the next time step. If the model chooses to predict, then the ground truth label will not be requested at the same time, and a $\vec{0}$ vector will be sent to the model along with the next instance instead of the true label.

r_t is the reward or penalty received after taken action a_t in state s_t, and γ represents the discount factor for future rewards. At each time step, once the model performs an action, one of the following three rewards is given: R_{cor} for correctly predicting the label, R_{inc} for incorrectly predicting the label, R_{req} for requesting the label. The goal is to

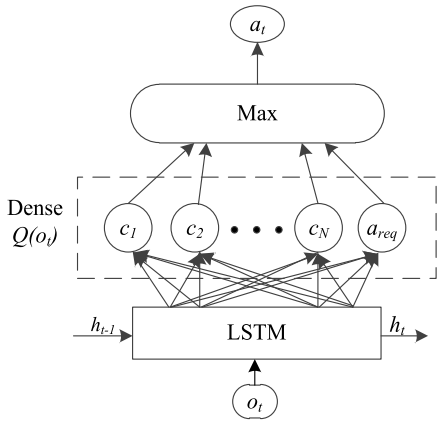


FIGURE 2. Schematic diagram of the proposed reinforcement one-shot active learning (ROAL).

maximize the sum of rewards received in this episode.

$$r_t = \begin{cases} R_{cor}, & \text{if predicting and } \hat{y}_t = y_t \\ R_{inc}, & \text{if predicting and } \hat{y}_t \neq y_t \\ R_{req}, & \text{if a label is requested} \end{cases} \tag{2}$$

B. METHODOLOGY

The purpose of reinforcement learning is to seek practical and superior strategies in complex control and prediction tasks through interaction with the environment. Through explorations and exploitation, it can learn from actions by receiving positive and negative reinforcements following the action performed. In this paper, an efficient model-free reinforcement learning method Q-learning is employed to learn an optimal policy $\pi^*(s_t)$ for maximizing the expected reward for any initial state. It can estimate the expected utility from the available operations and adapt to random transitions without understanding the system model [54], thus, Q-learning has been widely used in various decision-making problems [55]. In this paper, a long short-term memory (LSTM) is used to approximate the action-value function of Q learning and is connected to a fully-connected output layer to output the Q values, as depicted in Fig. 2.

In reinforcement learning, a definition of an objective function is required to show what action is good in the long term. The idea of Q-learning is not to require a model of the environment, but to optimize a Q function that can be directly calculated:

$$Q(s_t, a_t) = r_t + \gamma \max_{a_{t+1} \in \mathcal{A}} Q(s_{t+1}, a_{t+1}) \tag{3}$$

where γ is a discount factor between 0 and 1.

The policy which is taken at s_t is represented as $\pi(s_t)$, and outputs an action a_t at time t. The optimal policy $\pi^*(s_t)$ which is better than or equal to other policies always exists. $\pi^*(s_t)$ is the strategy that maximizes the optimal action-value function $Q^*(s_t, a_t)$. The action-values are consistently updated after observing rewards received after taking different actions in different states, and should ultimately result in a policy that is an estimate of the optimal policy π^*. Thus, the action which chosen by the model is given by the optimal policy π^* and can be calculated as:

$$a_t = \pi^*(s_t) = \underset{a_t \in \mathcal{A}}{\operatorname{argmax}}\, Q^*(s_t, a_t) \tag{4}$$

According to the Bellman equation, the optimal action-value function can be derived as:

$$Q^*(s_t, a_t) = \mathbb{E}_{s_{t+1}}[r_t + \gamma \max_{a_{t+1} \in \mathcal{A}} Q^*(s_{t+1}, a_{t+1})] \tag{5}$$

Normally, $Q(s_t, a_t)$ is represented by a function approximator and its parameters is optimized by minimizing the Bellman error. Woodward and Finn [20] derived the loss function as following:

$$L(\theta) := \sum_t \left[Q(o_t, a_t) - \left(r_t + \gamma \max_{a_{t+1} \in \mathcal{A}} Q^*(s_{t+1}, a_{t+1}) \right) \right]^2 \tag{6}$$

Here θ represents the model parameters, and o_t are the observations (instances) which the agent receives.

However, the loss function in Woodward's work [20] only considers the maximum value of Q. Thus, in the early stages of training, this loss function tends to be inefficient and prone to encounter gradient vanishing phenomenon. As an important concept in Shannon's information theory, cross entropy is mainly used to estimate the difference between two probability distributions and has been widely used in many machine learning methods to define a loss function. Intuitively, we want to introduce the cross-entropy term to the loss function to make the label prediction probability distribution output by the current model closer to the probability distribution of the real label [21], thus avoiding the shortcomings, speeding up the training and improving the efficiency of the model. The loss function we design is:

$$L(\theta) := \begin{cases} \sum_t \left(\begin{array}{l} \left[Q(o_t, a_t) - \left(r_t + \gamma \max_{a_{t+1} \in \mathcal{A}} Q^*(s_{t+1}, a_{t+1}) \right) \right]^2 \\ -p\left(Q(o_t, a_t)\right) \log(q\,(label\ (t))) \end{array} \right) \\ \qquad \text{if predicting} \\ \sum_t \left[Q(o_t, a_t) - \left(r_t + \gamma \max_{a_{t+1} \in \mathcal{A}} Q^*(s_{t+1}, a_{t+1}) \right) \right]^2 \\ \qquad \text{if requesting} \end{cases} \tag{7}$$

where $p(Q(o_t, a_t))$ are the probability distribution of $Q(o_t, a_t)$, $q\,(label\ (t))$ are the probability distribution of the true label at time step t.

A long short-term memory (LSTM) network [50] is used here, which is connected to a fully-connected layer to output the Q values. Each bit of the vector, which is the output of $Q(o_t)$, corresponds to an action:

$$Q(o_t, a_t) = Q(o_t) \cdot a_t \tag{8}$$

$$Q(o_t) = W^{hq} h_t + b^q \tag{9}$$

b^q is the bias vector of the action-value, h_t is the hidden state vector also known as output vector of the LSTM unit, W^{hq} are the weight metrics mapping from the LSTM output to action-values. The forms of the equations for the forward pass of an LSTM unit with a forget gate we used are:

$$\hat{g}^f, \hat{g}^i, \hat{g}^o, \hat{c}_t = W^o o_t + W^h h_{t-1} + b \quad (10)$$
$$g^f = \sigma(\hat{g}^f) \quad (11)$$
$$g^i = \sigma(\hat{g}^i) \quad (12)$$
$$g^o = \sigma(\hat{g}^o) \quad (13)$$
$$c_t = g^f \odot c_{t-1} + g^i \odot \tanh(\hat{c}_t) \quad (14)$$
$$h_t = g^o \odot \tanh(c_t) \quad (15)$$

Here, $\hat{g}^f, \hat{g}^i, \hat{g}^o$ respectively represent the forget gates, input gates, and output gates. Where $\hat{c}_t$ denotes the candidate cell state and c_t represents the new LSTM cell state. W^o denotes the weights mapping from the observation to the gates, and W^h represents the weights mapping from the hidden state to the candidate cell state. b denotes the bias vector. $\sigma(\cdot)$ denotes an element-wise sigmoid function. $\odot$ is element-wise product, and $\tanh(\cdot)$ represents the hyperbolic tangent function.

IV. EXPERIMENTS

Two classification tasks were examined using our proposed ROAL model under an active one-shot learning set-up, and the results of the ROAL model are compared with the results of previous studies.

A. AMSTERDAM LIBRARY OF OBJECT IMAGES

1) SETUP

We perform our first experiments on the Amsterdam Library of Object Images (ALOI) dataset [22] to show the general performance for target recognition. ALOI is a color image collection, consisting of 1000 classes of small objects, with 108 images of each object, giving 108,000 total instances. The dataset was split into 700 objects for training and keep the rest 300 objects for testing. Our model interacts with new objects it did not encounter in the training process to measure its test performance.

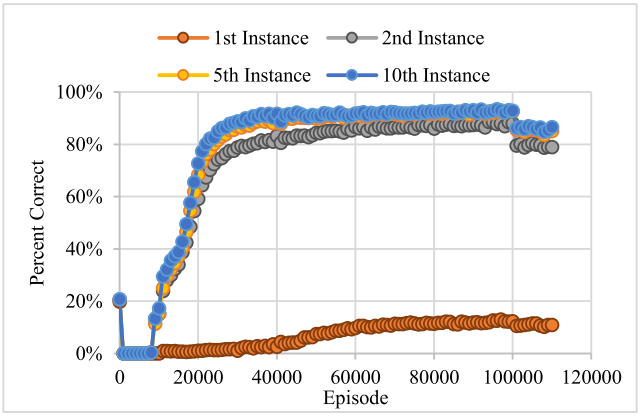
Following the episodical stream-based setup, every episode consists of a series of 50 images from the ALOI dataset. In each episode, these 50 instances were randomly selected from 5 different classes, and these classes were randomly drawn before every episode without replacement. Here, the number of instances from each class may be unbalanced. Each selected class in the episode wasn't labeled with their true label, but a pseudo-label randomly assigned when constructing the episode. The pseudo-labels are simply one-hot vector of size equal to the number of classes drawn, giving y_t. A single layer LSTM with 200 hidden units was used to represent Q. We used Adam with the default parameters [56] to optimize the weights of the model. A grid search was performed over the following hyper-parameters, and the hyper-parameters of the results reported in this article are listed as follows. During training process, the model employed an epsilon greedy exploration strategy, with $\epsilon = 0.05$. The discount factor γ was set to 0.5. Unless otherwise stated, each training step consisted of a batch of 100 episodes, the reward values were set as: $R_{cor} = +1$, $R_{inc} = -1$, and $R_{req} = -0.05$. The training was carried out on 100,000 episodes. For evaluation, 20 episodes were set as a group from the test set and the average accuracy, request, and precision rate were computed. And 10,000 episodes of evaluation were conducted after training.

2) RESULTS

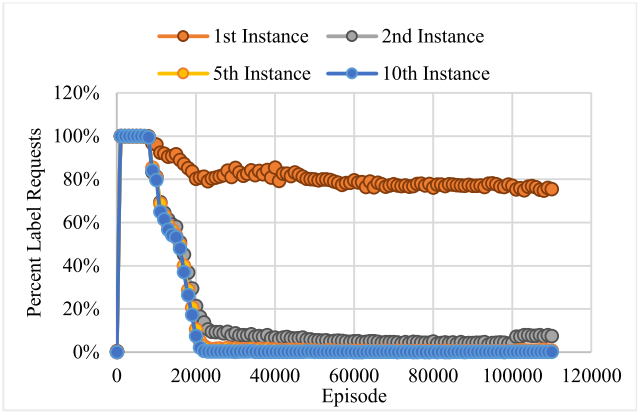
Here we represent two experimental results of our model on the ALOI dataset. In the first experiment, both active one-shot learning (AOL) [20] and ROAL model were tested on the task in Fig. 1 with the same parameters set-up. During training process, the 1st, 2nd, 5th, and 10th instances of all classes in each episode are identified. Notably, in this analysis, label requests are considered to be incorrect label predictions when calculating the accuracy. The models were trained on 100,000 episodes from the training set. After that, training was ceased, and the models evaluated on 10,000 more test episodes. In these episodes, no further update occurred, and then the model was run on never-before-seen classes pulled from a disjoint test set. We report the results in Fig. 3 and Fig. 4.

As can be seen from the figures, the ROAL we proposed learns to query the label for early instances of a class and makes more predictions for later instances. Meanwhile, the accuracy of the model is improved on subsequent instances of a class. Compared with AOL, ROAL converges faster with higher accuracy and lower request rate. ROAL introduces cross entropy into the loss function, which greatly speeds up the training, and saves time and computing resources.

Then, we performed another experiment to explore whether the model can effectively reason its own uncertainty. In previous experiments, instances in each episode were randomly arranged. In this experiment, in order to explore the model's action strategy, the order of instance was manually designed. Under the setting of this task, experiments were conducted on the trained model, and three test classes were randomly chosen for each episode. Two groups of experiments were carried out. In both groups, 1000 episodes were run without learning and the request percentage of episodes for each time step was recorded. In the first group, three instances were assigned which came from different classes to the model at the beginning of each episode. After that, three instances from different classes were given, respectively. We reported the label request rate for the first six time-steps in each episode separately. As can be seen in Fig. 5 (a), after the model saw an instance of that class, it should be able to recognize it next time it sees an instance of the same class, thus, the request rate for later instances of the same class was greatly reduced. This result is consistent with the original intention of active learning. If representative samples can be

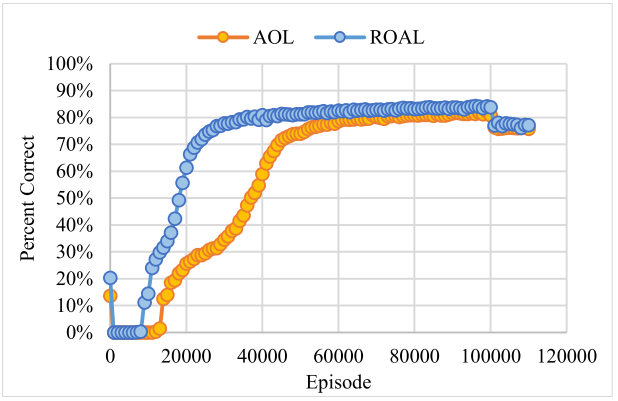


(a) Accuracy

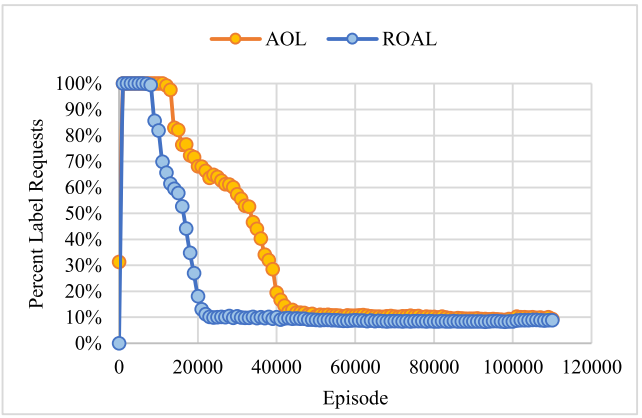


(b) Label requests

FIGURE 3. ROAL Accuracies (a) and label requests (b) per episode for the 1st, 2nd, 5th, and 10th instances of all classes.



(a) Accuracy



(b) Label requests

FIGURE 4. Comparison of overall accuracy (a) and request rate (b) results between ROAL and AOL.

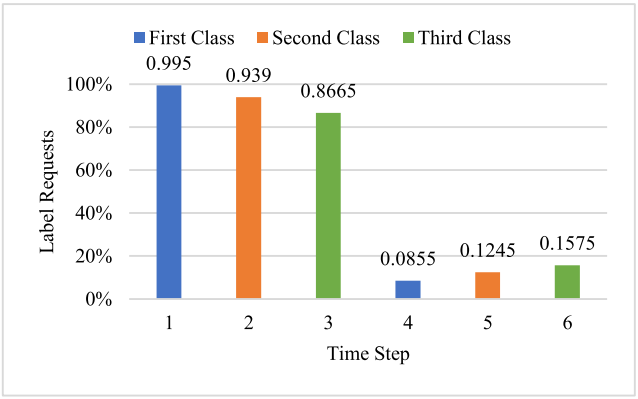
effectively selected for labeling, the cost of manual labeling can be greatly reduced. However, existing experiments have not been able to prove whether the model chooses actions based on uncertainty of instances, since a naive strategy is likely to be learned, which always requires labels in the first few steps. For further confirmation, another group of experiments was set as: two instances from the first class were given, followed by two instances from the second class and two instances from the third class. As shown in Fig. 5 (b), the label request rate of the second, fourth, and sixth time step are greatly reduced, and the label request rates of the third, and fifth time step are greatly increased. The difference in request rates between these time steps and the similarity between the percentages of label requests of all the first instance of each class indicate that the model chooses the action based on the uncertainty of instances, since the model is able to query the label when a new class appears and rapidly learn new concepts after that.

B. AIRCRAFT TYPE RECOGNITION

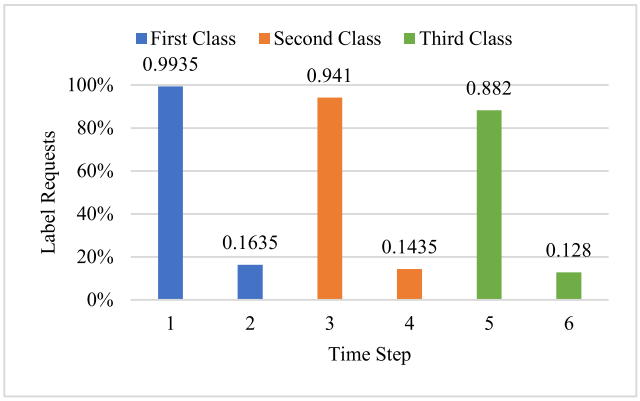
1) SETUP

The aircraft type classification dataset covers 215 classes of aircraft, with each class consisting of 20 aircraft, for a total of 4300 aircraft. It is based on the time-series data of a month's aircraft flight tracks collected by multiple sensors, and it contains the track information of both military and civilian aircraft. This form of flight track data can be passively collected from far away in almost any location, which varies from sound and radar data which are limited in location (both) and are active (radar) [34]. The flight data is comprised of irregular intervals that make up the record of each track. We extracted the motion features as the inputs of the model [1]. The dataset was split into 152 classes for training and kept the remaining 67 classes for testing.

For the first experiment, in each episode, a series of 30 aircraft tracks were randomly selected, these 30 instances were randomly selected from 3 different classes, and these classes were randomly drawn before every episode without replacement. The number changed to 50 or 70 tracks per episode when the number of classes per episode changed to 5 and 7. Q is represented by an LSTM with 600 units. We used Adam with the default parameters [56] to optimize the weights of the model. The following hyper-parameters were chosen by a grid search and are listed as follows. An epsilon-greedy exploration strategy with $\epsilon = 0.1$ was used for action selection. The discount factor γ was set to 0.6. In experiments on aircraft type recognition task, unless otherwise stated, the reward values were set as:

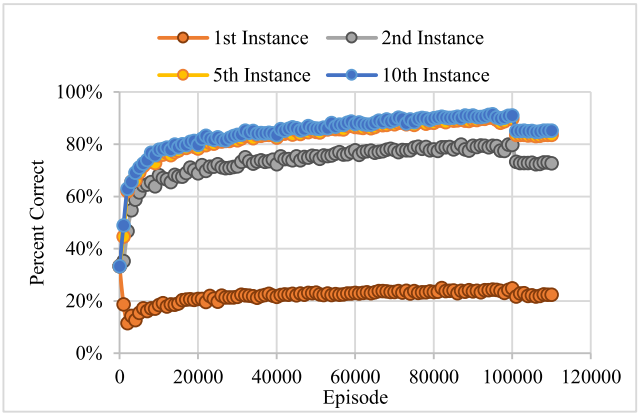


(a) Uncertainty test 1

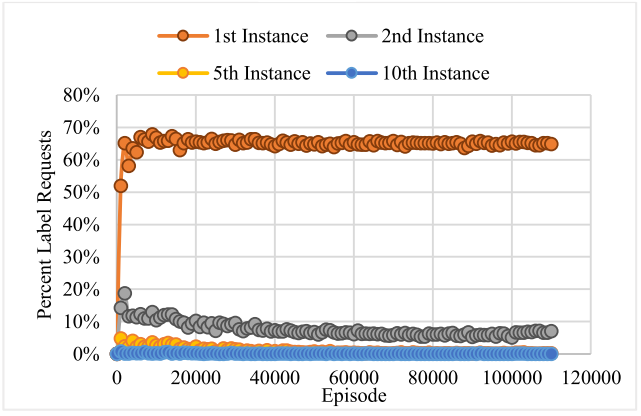


(b) Uncertainty test 2

FIGURE 5. Uncertainty test results.



(a) Accuracy



(b) Label requests

FIGURE 6. Label requests (a) and accuracies (b) per episode for the 1st, 2nd, 5th, and 10th instances of all classes.

$R_{cor} = +1$, $R_{inc} = -1$, and $R_{req} = -0.3$. The training was carried out on 100,000 episodes. For evaluation, 20 episodes were set as a group from the test set and the average accuracy, request, and precision rates were computed. And 10,000 episodes of evaluation were conducted after training.

2) FEATURE EXTRACTION

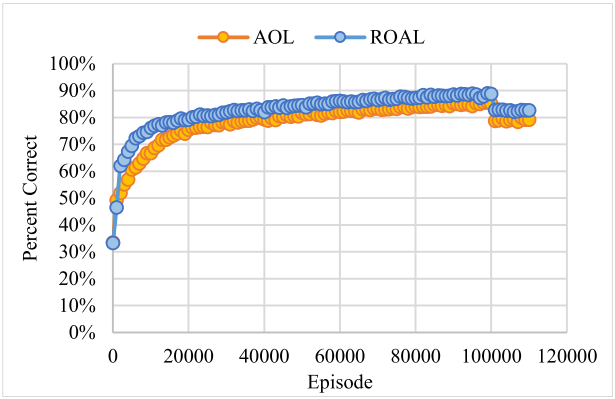
Because of the differences in aircraft performance and pilot flight habits, useful motion features such as maximum speed, cruising speed, maximum acceleration, maximum rate of climb were extracted as the input [1].

3) RESULTS

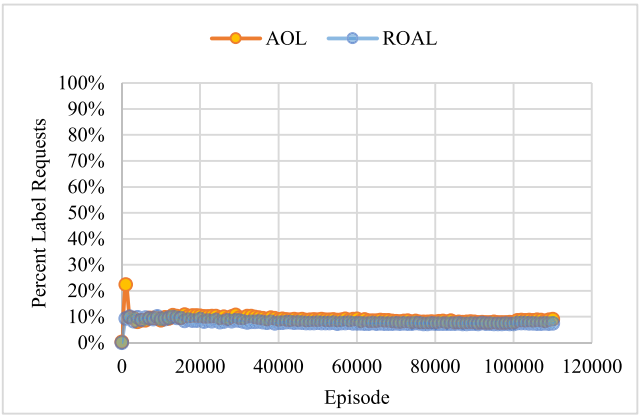
In Fig. 6 and Fig. 7 we report the results of our active model on aircraft type recognition task.

As shown in Fig. 6, since the ROAL model learns to query the label for early instances of each class, first-instance accuracy is poor. We can also conclude that ROAL leads to more label predictions for later instances according to the sharp drop in label request rates for later instances. At the same time, the prediction accuracy of the model is further improved on later instances of a class, close to 85%. As shown in Fig. 7, compared with AOL, ROAL converges faster and achieves higher accuracy. Since the tasks we show here are relatively simple, each episode contains only 3 different categories, the label request rate of AOL and ROAL are almost the same low. Student's paired t-test was conducted to evaluate the statistical significance of the comparison results for ROAL and AOL. When the p-value in the hypothesis test was less than 0.05, the result was considered significant. In our results, the statistical significance levels of both the training and test stages of accuracy are significantly lower than 0.05, indicating that the results of ROAL are significantly superior to the results of AOL. These data show that ROAL greatly speeds up the training, effectively avoids the inefficiency in the early training stage, and saves considerable time and computing resources by introducing cross entropy into the loss function.

In order to further compare ROAL and AOL, Fig. 8 shows the receiver operating characteristic (ROC) curve analyses results in the multiclassification task. The ROC curve is a graphical plot of the true positive rate (TPR) against the false positive rate (FPR) as its discrimination threshold is varied. It can clearly illustrate the diagnostic ability of a classifier system. A ROC plane is defined by FPR as the X-axis and TPR as the Y-axis, respectively, the axes range from 0 to 1. A random guess would give a diagonal dotted straight line connecting (0,0) to (1,1). The diagonal divides the ROC space. Any classifier that appears above the diagonal



(a) Prediction accuracy



(b) Label requests

FIGURE 7. Comparison of overall accuracy (a) and request rate (b) results between ROAL and AOL.

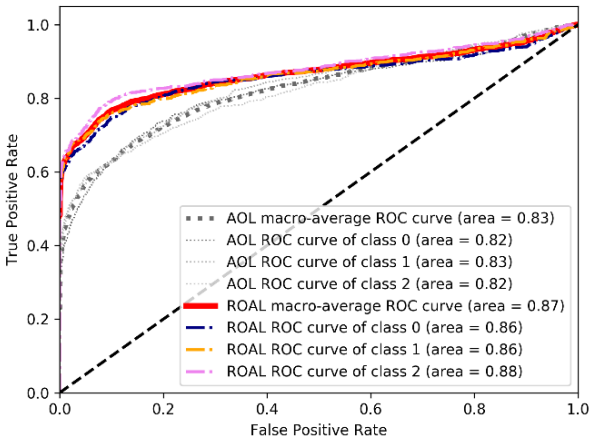


FIGURE 8. ROC plot with AUC values for AOL and ROAL.

performs better than random guessing, whereas curves below the line represent worse classification performances. Since we study the case of multiclassification, not only the ROC curves of the two algorithms for each class but also the macro-average ROC curves that reflect the overall classification effect of the two algorithms are presented. As can be seen in Fig. 8, the ROAL method achieves better upper-left ROC curve results than the AOL method.

TABLE 1. Test set classification accuracies and the percentage of label requests per episode.

%	AOL		ROAL	
	Accuracy	Requests	Accuracy	Requests
Accuracy	72.25	8.846	**76.33**	**7.504**
Prediction	79.25	8.846	**82.52**	**7.504**
$R_{inc} = -2$ prediction	90.07	29.73	**90.23**	**19.4**
$R_{inc} = -3$ prediction	94.06	42.13	**95.07**	**32.76**
$R_{inc} = -4$ prediction	96.31	49.46	**96.89**	**43.77**
$R_{inc} = -5$ prediction	97.29	55.09	**97.82**	**49.13**

The areas under the curve (AUCs) of the ROC plot were often used for model comparison in machine learning. The AUC can be calculated by accumulating the trapezoidal areas between each ROC point. The AUC value lies between 0 and 1, and the higher AUC value, the better classification performance. As can be seen in Fig. 8, the macro average AUC of ROAL is higher, which is 0.87, while the macro average AUC of the AOL method is 0.83. And the AUC values for each class of ROAL is also higher than AOL. The results of ROC-AUC analyses show that, compared with the AOL, the ROAL algorithm effectively improves the classification performance.

It is a natural idea to increase the penalty for misprediction to improve the accuracy of the model. And prediction accuracy is the most important thing in aircraft recognition task. In reinforcement learning, this goal can be achieved by changing the setting of reward function. To explore the impact of this, we further trained models using different reward values, which are $R_{inc} = -1$, $R_{inc} = -2$, $R_{inc} = -3$, $R_{inc} = -4$, and $R_{inc} = -5$. At the same time, we show the results of the AOL model presented on the same problem. As shown in TABLE 1, the prediction accuracy increases with the increase of the penalty of incorrect labeling. Compared to AOL, ROAL achieves higher accuracy and a lower request rate with the same reward value setting. The experimental results also verified that the ROAL model can make trade-offs between high prediction accuracy of numerous label requests and a small number of label requests with low prediction accuracy. Higher prediction accuracy can be achieved by increasing the penalty value for wrongly predicting labels. Previous state-of-the-art aircraft recognition studies have established a baseline of over 90% recognition accuracy. As R_{inc} becomes more negative, ROAL approaches the accuracy over 97%, with less than 50% label request rate. Notably, we can conclude from the table that with the increase of model accuracy, the request rate increases rapidly. When the model accuracy exceeds more than 95%, the cost of increasing 1% accuracy is the increment of more than 11% label request rate. Therefore, properly setting the reward value function poses a vital impact on the learning effect of the model.

TABLE 2. Results for various architectures on the aircraft recognition task.

%		3 classes per episode	5 classes per episode	7 classes per episode
Random	Accuracy	62.63	50.04	43.82
	Request	80	80	80
Density	Accuracy	62.63	50.32	43.99
	Request	80	78	77.14
LAL	Accuracy	62.63	50.16	43.82
	Request	80	78	80
Unc	Accuracy	65.21	50.56	44.09
	Request	60	70	77.14
QBC	Accuracy	64.88	52.15	45.06
	Request	66.67	68	67.14
Supervised	Accuracy	78.2	66.2	59.5
	Request	100	100	100
AOL $R_{inc} = -1$	Accuracy	79.25	72.49	73.28
	Request	8.846	26.06	73.2
AOL $R_{inc} = -2$	Accuracy	90.07	89.11	87.39
	Request	29.73	49.88	69.68
ROAL $R_{inc} = -1$	Accuracy	82.52	74.5	77.09
	Request	**7.504**	**18.24**	**33.5**
ROAL $R_{inc} = -2$	Accuracy	**90.23**	**90.64**	**92.74**
	Request	19.4	44.7	61.67

The experiments were further expanded by increasing the number of classes per episode. In the same task, the ROAL model was compared to AOL, a supervised learning model and 5 active learning methods [57] (Random Sampling (Random) [58], Diversity promoting sampling (Density) [44], Learning Active Learning (LAL) [45], Uncertainty sampling (Unc) [41], Query By Committee (QBC) [43]) in the same task, where the model must deal with never-before-seen classes in the test set. The results are shown in TABLE 2, and the rewards for AOL and ROAL were set as: $R_{cor} = +1$, and $R_{req} = -0.3$. For active learning methods, one labeled instance for each class was needed for setup at the beginning of each episode. The loss of the supervised learning model is the cross entropy between the true label and the predicted label, and the true label is always presented in the subsequent time step. For consistency, we used the same LSTM model in this supervised task [49], and the softmax modification is performed on the output without extra bits for the "request label" action. The results show that the traditional supervised learning method and active learning methods cannot rapidly learn new concepts, so they may be incapable of the task of recognizing new targets in one-shot learning. Through the increment of the number of classes per episode, the ability of the ROAL algorithm to handle more complex tasks is further demonstrated. At the same time, compared with others, the ROAL model significantly reduces the number of requests for labels while achieving the same or even higher accuracy. However, we also found that as the complexity of the problem increases, the request rate of the label also increases rapidly, and the excessive label request rate means a large consumption of human resources. So, in the face of more complex issues, LSTM-based networks will no longer be competent, and a more powerful one-shot learning approach should be introduced. Notably, as explained in [49], human performance is a relevant baseline for one-shot learning. However, a central memory store is limited to 3 to 5 meaningful items in young adults [59]. Therefore, for the task like aircraft type recognition with the number of classes far beyond 5, this type of binding surpasses human working memory capacity, which is limited to storing only a handful of arbitrary bindings [49].

Compared to our previous work using supervised learning methods for aircraft type recognition [1], methods based on reinforcement one-shot active learning can significantly reduce the dependence on label data and achieve the same or even better model accuracy.

V. CONCLUSION

As an essential technology in air traffic management, aircraft type recognition is attracting increasing amounts of attention from scholars. The existing studies have been mostly based on supervised graphic image processing, which is inherently deficient in highly dynamic real-time applications. In this paper, we first develop a model that learns actively via reinforcement learning with a label query strategy based on data characteristics. Secondly, we apply this meta active one-shot learning approach to target recognition tasks using ALOI and aircraft type recognition datasets. The experimental results demonstrate that the model is good at rapidly learning new concepts and can transform an engineering heuristic selection of samples into learning strategies based on data. Compared to previous studies, we significantly accelerate the convergence, improve the stability, decrease the number of label requests and improve the accuracy of the model. Notably, the proposed model can learn when to label examples and when to request a label instead; thus, it meets the need of intelligent air traffic management and has a wide range of applications.

In future work, we plan to evaluate our approach on more complex datasets and expand the scope of the study to a wider range of targets. For this, we may need a more sophisticated one-shot learning approach such as Matching Network [15] or Memory-Augmented Neural Networks [49].

REFERENCES

[1] H. Huang, J. Huang, Y. Feng, Z. Liu, T. Wang, L. Chen, and Y. Zhou, "Aircraft type recognition based on target track," *J. Phys., Conf. Ser.*, vol. 1061, no. 1, 2018, Art. no. 012015.

[2] P. Bachman, A. Sordoni, and A. Trischler, "Learning algorithms for active learning," 2017, *arXiv:1708.00088*. [Online]. Available: https://arxiv.org/abs/1708.00088

[3] M. W. Huijser and J. C. van Gemert, "Active decision boundary annotation with deep generative models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5296–5305.

[4] G. Contardo, L. Denoyer, and T. Artieres, "A meta-learning approach to one-step active-learning," 2017, *arXiv:1706.08334*. [Online]. Available: https://arxiv.org/abs/1706.08334v2

[5] X. Zhou, *Semi-Supervised Learning Literature Survey, Computer Sciences, University of Wisconsin-Madison*. Accessed: Dec. 14, 2007. [Online]. Available: http://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf

[6] Q. Wang, X. Zhao, J. Huang, Y. Feng, Z. Liu, J. Su, Z. Luo, and G. Cheng. *Addressing Complexities of Machine Learning in Big Data: Principles, Trends and Challenges from Systematical Perspectives*. Accessed: Oct. 12, 2017. [Online]. Available: https://www.preprints.org/manuscript/201710.0076/v1

[7] A. Padmakumar, P. Stone, and R. J. Mooney, "Learning a policy for opportunistic active learning," in *Empirical Methods Natural Language Processing*, Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 1347–1357.

[8] A. Tian and M. Lease, "Active learning to maximize accuracy vs. effort in interactive information retrieval," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Beijing, China, Jul. 2011, pp. 145–154.

[9] D. Yu, B. Varadarajan, L. Deng, and A. Acero, "Active learning and semi-supervised learning for speech recognition: A unified framework using the global entropy reduction maximization criterion," *Comput. Speech, Lang.*, vol. 24, no. 3, pp. 433–444, 2010.

[10] H. Rong, B. M. Namee, and S. J. Delany, "Active learning for text classification with reusability," *Expert Syst. Appl.*, vol. 45, pp. 438–449, Mar. 2016.

[11] M. Davy and S. Luz, "Dimensionality reduction for active learning with nearest neighbour classifier in text categorisation problems," in *Proc. Int. Conf. Mach. Learn. Appl.*, Cincinnati, OH, USA, Dec. 2007, pp. 292–297.

[12] G. V. Cormack and M. R. Grossman, "Scalability of continuous active learning for reliable high-recall text classification," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Indianapolis, IN, USA, Oct. 2016, pp. 1039–1048.

[13] E. Kriminger, J. Cobb, and J. Principe, "Online active learning for automatic target recognition," *IEEE J. Ocean. Eng.*, vol. 40, no. 3, pp. 583–591, Jul. 2015.

[14] B. M. Lake, R. Salakhutdinov, and J. B. Tenebaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.

[15] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," 2016, *arXiv:1606.04080*. [Online]. Available: https://arxiv.org/abs/1606.04080

[16] A. Puzanov and K. Cohen, "Deep reinforcement one-shot learning for artificially intelligent classification systems," 2018, *arXiv:1808.01527*. [Online]. Available: https://arxiv.org/abs/1808.01527

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[18] R. S. Sutton and A. G. Barto, "Introduction," in *Reinforcement Learning: An Introduction*, vol. 1, 2nd ed. Cambridge, MA, USA: A Bradford Book, 2018, pp. 1–4. [Online]. Available: http://incompleteideas.net/book/RLbook2018.pdf

[19] K. Pang, M. Dong, Y. Wu, and T. Hospedales, "Meta-learning transferable active learning policies by deep reinforcement learning," 2018, arXiv:1806.04798. [Online]. Available: https://arxiv.org/abs/1806.04798

[20] M. Woodward and C. Finn, "Active one-shot learning," 2017, *arXiv:1702.06559*. [Online]. Available: https://arxiv.org/abs/1702.06559

[21] H. Huang, J. Huang, Y. Feng, Z. Liu, Q. Wang, J. Zhang, and L. Chen, "On the improvement of reinforcement active learning with the involvement of cross entropy to address one-shot learning problem," *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0217408.

[22] J.-M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," *Int. J. Comput. Vis.*, vol. 61, no. 1, pp. 103–112, 2005.

[23] Y. Feng, X. Yang, and G. Cheng, "Stability in mean for multi-dimensional uncertain differential equation," *Soft Comput.*, vol. 22, no. 17, pp. 5783–5789, Sep. 2018.

[24] B. Kamgar-Parsi, B. Kamgar-Parsi, and A. K. Jain, "Automatic aircraft recognition: Toward using human similarity measure in a recognition system," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Fort Collins, CO, USA, Jun. 1999, pp. 268–273.

[25] J.-W. Hsieh, J.-M. Chen, C.-H. Chuang, and K.-C. Fan, "Aircraft type recognition in satellite images," *IEE Proc. Vis., Image Signal Process.*, vol. 152, no. 3, pp. 307–315, Jun. 2005.

[26] Y. Liu, "Aircraft type recognition based on convex hull features and SVM," *Proc. SPIE*, vol. 6786, Nov. 2007, Art. no. 67863Q.

[27] A. A. Somaie, A. Badr, and T. Salah, "Aircraft recognition system using eigenvector technique," in *Proc. 16th Nat. Radio Sci. Conf.*, Cairo, Egypt, Feb. 1999, pp. C29/1–C29/9.

[28] J. Zuo, G. Xu, K. Fu, X. Sun, and H. Sun, "Aircraft type recognition based on segmentation with deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 282–286, Feb. 2018.

[29] P. Joris Zwart, "Aircraft recognition from features extracted from measured and simulated radar range profiles," Ph.D. dissertation, Dept. Sci., Universiteit Van Amsterdam, The Netherlands, 2003. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.7822

[30] X.-D. Li, J.-D. Pan, and J. Dezert, "A target recognition algorithm for sequential aircraft based on DSmT and HMM," *Acta Automatica Sinica*, vol. 40, no. 12, pp. 2862–2876, 2014.

[31] H. J. Rong, Y. X. Jia, and G. S. Zhao, "Aircraft recognition using modular extreme learning machine," *Neurocomputing*, vol. 128, no. 5, pp. 166–174, Mar. 2014.

[32] S. A. Dudani, K. J. Breeding, and R. B. Mcghee, "Aircraft identification by moment invariants," *IEEE Trans. Comput.*, vol. C-26, no. 1, pp. 39–46, Jan. 1977.

[33] T. P. Wallace and P. A. Wintz, "An efficient three-dimensional aircraft recognition algorithm using normalized Fourier descriptors," *Comput. Graph. Image Process.*, vol. 13, no. 2, pp. 99–126, 1980.

[34] S.-J. Huang, J.-L. Chen, X. Mu, and Z.-H. Zhou, "Cost-effective active learning from diverse labelers," in *Proc. Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 1879–1885.

[35] B. Settles, "Active learning literature survey," Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. 1648, 2010. [Online]. Available: http://burrsettles.com/pub/settles.activelearning.pdf

[36] P. Resnick and H. R. Varian, "Recommender systems," *Commun. ACM*, vol. 40, no. 3, pp. 56–59, Mar. 1997.

[37] N. Houlsby, J. M. Hernández-Lobato, and Z. Ghahramani, "Cold-start active learning with robust ordinal matrix factorization," in *Proc. 31st Int. Conf. Int. Conf. Mach. Learn.*, Beijing, China, vol. 2014, pp. 766–774.

[38] M. Sun, F. Li, J. Lee, K. Zhou, G. Lebanon, and H. Zha, "Learning multiple-question decision trees for cold-start recommendation," in *Proc. 6th ACM Int. Conf. Web Search Data Mining*, Rome, Italy, Feb. 2013, pp. 445–454.

[39] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Batch mode active learning and its application to medical image classification," in *Proc. 23rd Int. Conf. Mach. Learn.*, Pittsburgh, PA, USA, Jun. 2006, pp. 417–424.

[40] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Stream-based active learning for sentiment analysis in the financial domain," *Inf. Sci.*, vol. 285, pp. 181–203, Nov. 2014.

[41] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *Proc. 17th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Dublin, Ireland, Jul. 1994, pp. 3–12.

[42] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, Pittsburgh, PA, USA, Jul. 1992, pp. 287–294.

[43] N. Abe and H. Mamitsuka, "Query learning strategies using boosting and bagging," in *Proc. 15th Int. Conf. Mach. Learn.*, Madison, WI, USA, Jul. 1998, pp. 1–9.

[44] S. Ebert, M. Fritz, and B. Schiele, "RALF: A reinforced active learning formulation for object class recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 3626–3633.

[45] K. Konyushkova and S. Raphael, "Learning active learning from data," in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 4228–4238.

[46] E. Lughofer, "Single-pass active learning with conflict and ignorance," *Evolving Syst.*, vol. 3, no. 4, pp. 251–271, Dec. 2012.

[47] W. Chu, M. Zinkevich, L. Li, A. Thomas, and B. Tseng, "Unbiased online active learning in data streams," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Diego, CA, USA, Aug. 2011, pp. 195–203.

[48] Y. Feng, L. Dai, J. Gao, and G. Cheng, "Uncertain pursuit evasion game," *Soft Comput.*, Dec. 2018. doi: 10.1007/s00500-018-03689-3.

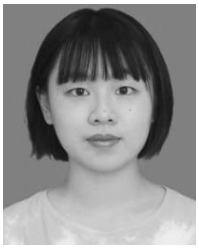
[49] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "One-shot learning with memory-augmented neural networks," 2016, *arXiv:1605.06065*. [Online]. Available: https://arxiv.org/abs/1605.06065

[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[51] J. Shore and R. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 1, pp. 26–37, Jan. 1980.

[52] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 705–712, 1996.

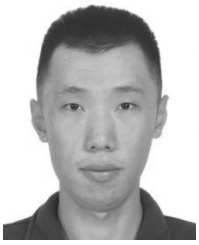
[53] B. Settles and M. Craven, "An analysis of active learning strategies for sequence labeling tasks," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Honolulu, HI, USA, Oct. 2008, pp. 1070–1079.
[54] C. J. Watkins and P. Dayan, "Technical note: Q-learning," *Mach. Learn.*, vol. 8, no. 3–4, pp. 279–292, May 1992.
[55] W. Liu, Y. Tan, and Q. Qiu, "Enhanced Q-learning algorithm for dynamic power management with performance constraint," in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, Dresden, Germany, Mar. 2010, pp. 602–605.
[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2015, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[57] Y.-P. Tang, G.-X. Li, and S.-J. Huang. (2019). *ALiPy: Active Learning in Python*. [Online]. Available: http://parnec.nuaa.edu.cn/huangsj/alipy/index.html
[58] Z. Wang and J. Ye, "Querying discriminative and representative samples for batch mode active learning," *ACM Trans. Knowl. Discov. Data*, vol. 9, no. 3, pp. 1–23, 2015.
[59] N. Cowan, "The magical mystery four: How is working memory capacity limited, and why?" *Current Directions Psychol. Sci.*, vol. 19, no. 1, pp. 51–57, Feb. 2010.



HONGLAN HUANG was born in Hefei, China, in 1995. She received the B.S. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2017. She is currently pursuing the M.S. degree in management science and engineering with the National University of Defense Technology, Changsha, China. Her research interests include active learning, reinforcement learning, and one-shot learning.

YANGHE FENG is currently an Associate Research Fellow with the National University of Defense Technology. His research interests include human factors' engineering, cognitive computing, deep learning, deep reinforcement learning, and active learning.

JINCAI HUANG is currently an Associate Research Fellow with the National University of Defense Technology, Changsha, Hunan, China, and a Researcher with the Science and Technology on Information Systems Engineering Laboratory. His main research interests include artificial general intelligence, deep reinforcement learning, and multiagent systems.



JIARUI ZHANG received the B.S. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2015, and the M.S. degree in aeronautical and astronautical science and technology from the National University of Defense Technology, Changsha, China, where he is currently pursuing the Ph.D. degree in aeronautical and astronautical science and technology.

LI CHEN received the B.S. degree from the Department of Measurement and Control Technology, Lanzhou University of Technology, and the master's degree from the School of Information Science and Engineering, Lanzhou University, Lanzhou, China. She is currently pursuing the Ph.D. degree from the School of Systems Engineering, National University of Defense Technology, Changsha, China.