Multi-Source Data Fusion and Target Tracking of Heterogeneous Network Based on Data Mining

Xunzhong Quan[1,2*], Jie Chen[3]

[1] College of Electronic Engineering, Huainan Normal University, Huainan 232038, China
[2] College of Mechanical & Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[3] Huainan science and Technology Development Research Center, Huainan 232001, China

Corresponding Author Email: robertqxz700720@126.com

ABSTRACT

Thanks to the technical development of target tracking, the multi-source data fusion and target tracking has become a hotspot in the research of huge heterogenous networks. Based on millimeter wave heterogeneous network, this paper constructs a multi-source data fusion and target tracking model. The core of the model is the data mining deep Q network (DM-DQN). Through image filling, the length of the input vector (time window) was extended from 25 to 31, with the aid of CNN heterogeneous network technology. This is to keep the length of input vector in line with that of output vector, and retain the time features of eye tracking data to the greatest extent, thereby expanding the recognition range. Experimental results show that the proposed model achieved a modified mean error of only 1.5m with a tracking time of 160s, that is, the tracking effect is ideal. That is why the DM-DQN outperformed other algorithms in total user delay. The algorithm can improve the energy efficiency of the network, while ensuring the quality of service of the user. In the first 50 iterations, DM-DQN worked poorer than structured data mining. After 50 iterations, DM-DQN began to learn the merits of the latter. After 100 iterations, both DM-DQN and structured data mining tended to be stable, and the former had the better performance. Compared with typical structured data mining, the proposed DM-DQN not only converges fast, but also boasts a relatively good performance.

1. INTRODUCTION

The depth feature extracted from heterogenous network makes up for the shortcomings of traditional methods, providing a suitable tool for sketch research. Based on the depth feature, it is possible to realize the end-to-end retrieval system, which effectively improves the performance of target tracking. The merits of target tracking and line of sight (LOS) tracking could be combined through data fusion.

Yang [1] was the first to apply triple network to target tracking, and successfully improved the retrieval accuracy. Since then, triple network became the most popular modeling framework. Chen and Han [2] expanded the Sketchy Dataset of 60,502 images from ImageNet, and selected the sketch extended model to verify the target tracking model for heterogenous networks. Shu et al. [3] applied the twin network to coarse-grained sketch-based image retrieval (SBIR) problem, and proposed a convolutional neural network (CNN) for target features. Unlike the traditional manual method of feature extraction, this CNN solves the domain shift problem from a new perspective, and guides model training with the comparative loss function of CNN learning. Vo and Ma [4] introduced triple network to coarse-grained target tracking, compared several triple heterogeneous CNNs, and developed a novel zero-shot SBIR (ZS-SBIR) model. Further, external pixel information was embedded to the model, and two new loss functions were adopted to track visible and invisible targets. Vo et al. [5] conducted comparative experiments with

different weight sharing strategies between the anchor branch and the other two branches, namely, no-weight sharing, half-weight sharing, and full-weight sharing, and demonstrated that the half-weight sharing architecture can enhance the ability of class generalization, and achieve a map value 18% greater than twin network.

Using VGG-19 pretrained on ImageNet data, Li et al. [6] solved the problem that depth network requires lots of training data, and extracted the natural contour of images with the Candy operator. Compared with other edge extractors, Litc's algorithm can retain many textures, and enable the network to receive lots of information on pixels. Chen et al. [7] introduced multi-step object painting to the SBI, and decomposed the hand-drawn object and image edge graph into three visual representation layers, according to the time sequence information and the painting order. These three layers correspond to each other in the same layer. Ren et al. [8] proposed a new benchmark of ZS-SBIR, and designed conditional variant automatic encoder (CVAE) framework and counter automatic encoder framework (CAAE) based on VGGNet. Li et al. [9] took the target and image data as experimental inputs, and observed that not all regions of the image or target are key effective information for cross modal mapping in the generalization problem. Li et al. [10] pioneered the application of zero-shot learning (ZSL) and cross domain hash to SBIR, and greatly expanded the research scope of ZS-SBIR. Experimental results show that Li's method only achieved a 1% higher map than traditional methods, but the

introduction of the heterogeneous network is a breakthrough.

The above research mainly focuses on target tracking in structured networks. However, the above methods are too complex and inaccurate facing the current market demand, failing to meet the required energy efficiency of image tracking. Therefore, it is necessary to probe deep into the data fusion and target tracking of multi-source heterogenous network.

Based on millimeter wave heterogeneous network, this paper constructs a multi-source data fusion and target tracking model. The core of the model is data mining deep Q network (DM-DQN). Through image filling, the length of the input vector (time window) was extended from 25 to 31, with the aid of CNN heterogeneous network technology. This is to keep the length of input vector in line with that of output vector, and retain the time features of eye tracking data to the greatest extent, thereby expanding the recognition range. Further, 5G network technology was integrated to enhance the recognition efficiency and fluency of multi-source data fusion and target tracking of the system. Finally, DM-DQN was compared with structured data mining, and the following properties of the algorithm were tested: network density, link state, stability, and iterative relationship.

2. DATA MINING AND MULTI-SOURCE DATA FUSION AND TARGET TRACKING OF HETEROGENEOUS NETWORK

2.1 Target tracking of heterogenous network

Traditionally, researchers of target tracking regarded sketch as the expression of shape contour, and tried to describe sketch features with geometric relationship [11]. Features are usually extracted with the feature descriptors specially designed for sketches, e.g., edge direction histogram of binary shape feature descriptor, key shape learning, gradient field as natural image feature descriptor, and scale invariant feature. Then, the similarity between the grass image and the edge image is measured by Euclidean distance and other metrics, and the candidate image is selected through similarity matching. Finally, the output is sorted and retrieved, completing the image retrieval process. However, the sketch is so abstract that the generated feature descriptors cannot effectively fit the content of the sketch, and thus cannot be easily used in real scenes [12]. In addition, it is difficult to realize an end-to-end retrieval system by traditional methods, causing a substantial growth in workload.

In 2012, Hinton's team won an image recognition competition with the AlexNet, which set off a new wave of heterogeneous networks [13]. Heterogenous networks like CNN and deep neural network (DNN) become the primary tools to solve relevant problems [14]. Unlike traditional manual approach of feature extraction, heterogenous networks adopt a layer-by-layer design. Mimicking human perception, a heterogenous network can learn depth features of different levels (low, medium, or high levels) from hand-drawn sketches or natural images, and learn the abstract pixel information hidden in the target image [15]. In many fields, the fusion algorithm based on heterogeneous network has achieved good results. Studies have shown the immense value of using artificial neural network (ANN) fusion to handle nonlinear problems. For instance, ANN has an outstanding performance in obstacle detection [16].

2.2 Data fusion of bidirectional long short-term memory (BLSTM)

As an improved version of recurrent neural network (RNN), BLSTM overcomes the difficulty in processing long-term dependent information, and effectively acquires the context of temporal data [17]. Ilić et al. [18] proposed a deep CNN (DCNN) fusion framework, which fuses input features excellently layer by layer, but ignores the time features of data. To improve the fusion performance, two-layer CNN was combined with the LSTM, and single-layer CNN with BLSTM.

The coupling of DCNN with BLSTM can extract the temporal and spatial information of eye tracking data. The literature shows that, when the time window is longer than 1s, eye movement can be captured well. Since the sampling frame rate of the test data is between 24 and 25 frame per second (fps), the time window was set to 25. Specifically, each convolution layer performs a convolution operation, and the window length is reduced by 2. Thus, 6 data need to be filled for the three convolution layers.

Image filling is a strategy involving the followings steps: Read the input image, decode the JPEG content into RGB grid pixels, convert the results into floating-point tensor, and rescale the pixel values from (0, 255) to [0, 1]. Then, multiple images could be imported, and fused into a new image of the same class. The new image contains the information of the multiple input images. After that, linear interpolation is carried out in the potential space, and image fusion is conducted by a certain proportion. Finally, the length of input vector (time window) is extended from 25 to 31 through image filling, such as to match the length between input and output vectors, and retain the temporal features of eye tracking data [19].

For eye-CNN-BLSTM, the parameters are initialized empirically, and optimized through repeated experiments. The optimal scheme is as follows: First, one DCNN is adopted to extract the spatial information from the input data. The three convolution layers have 16, 8, and 4 convolution kernels, respectively. Before each convolution, the data are subjected to batch normalization (BN), because the data distribution can be changed through network training [20]. The convolution part adopts the rectified linear unit (ReLU) as the activation function. This sparse activation function can speed up network training [21]. Besides, the fully connected layer is adopted to connect all the features extracted from the convolution layer, and a dropout operation is added to enhance the generalization ability of the model [22]. Furthermore, the tanh function is taken as the activation function, to prevent the excessively large output of ReLU [23]. Finally, a fully connected layer is selected as the output layer of the fusion value.

2.3 Heterogenous network model for target tracking

Ground base station (G-BS) and unmanned aerial vehicle (UAV)-mounted base station (U-BS) are two millimeter wave heterogenous networks built for real hot scenes. The G-BS and U-BS are composed of macro- and micro- cell, respectively [24].

In order to reduce the traffic load of G-BS, it is assumed that U-BSs in the air are clustered and projected around G-BS. On the Euclidean plane, the position of a G-BS is modeled as a dense, homogeneous network of G-BSs, while that of a U-BS is modeled as a PCP with parent point process of G-BS, for the U-BS is deployed in hot spots. Therefore, the U-BS is further modeled as U-BS/G-BS, and its projection is scattered around

the G-BS, following the symmetrical Gaussian distribution with a variance of 2 [25].

A multi-relay multi-user millimeter wave hybrid heterogeneous network consists of a macro cell source node s, K small cell relay nodes $R_k(k = 1, 2, 3,..., K)$, and M ground destination nodes D_M ($M = 1, 2, 3,..., m$), which correspond to each relay node. Each relay node has a receiving antenna (RN) and a transmitting antenna (TN). Each other node has a single RN and a single TN. Each relay node is far from other relay nodes, and the M destination nodes of each RN form a cluster. Therefore, the following mathematical expression can be derived:

$$G\text{-}BS(x) = 2n\ln(\sigma) + n\ln(2\pi)$$
$$+ n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\} * DV \tag{1}$$

where, x is position; s is density. Moreover, we have:

$$y_i = IR*\beta(u_i, v_i) + \sum_{j=1}^{p}\beta_j(u_i, v_i)x_{ij} + \varepsilon_j\beta_j \tag{2}$$

where, IR is the trade-off coefficient of the objective function; DV is a soft max cross entropy loss function. In addition, we have:

$$\sigma_{ikjl} = \begin{cases} \frac{n}{\Delta_{ikjl}}\sqrt{\sum_{s=1}^{n}(x_{ik}(\varepsilon) - x_{jl}(\varepsilon))^2\Delta_{ikjl}(\varepsilon)} & \Delta_{ikjl} > 0 \\ 0 & \Delta_{ikjl} < 0 \end{cases} \tag{3}$$

$$y_i = f\left(\sum_{j=i}^{k}\omega_{ij}y_j - \theta_i\right) - C*N, i \neq j \tag{4}$$

where, n is the total number of initial training samples in X; K is the total number of occlusion transformations in each x; C is the total class. After the above steps, the relaxed nonconvex minimization problem can be written as:

$$L_r = \|D_I - D_O\|_2 \tag{5}$$

$$IR = \sum_{S-1}^{U}\sum_{d-1}^{K}f_s, DV_s, d \tag{6}$$

$$W(T) = K\left(y(T-1),....,y(t-n),u(T-d-1),...,u(k-d-n)\right) \tag{7}$$

$$W(T) = K\left(y(T-1), u(T-d-1)\right) \tag{8}$$

In addition, two additional layers can be constructed: one layer of one U-BS, and three layers of three G-BSs. The two layers are composed of inter-cluster U-BS and inter-cluster G-BS, respectively. Therefore, the network model can be extended as follows: Layer 0 of intra-cluster U-BS, Layer 1 of inter-cluster U-BS, layer 2 of intra-cluster G-BS, and layer 3 of inter-cluster G-BS. Then, the logarithmic hyperbolic cosine function can be compared with other functions:

$$U = \sum_{i=1}^{g}\left\{P_i\left|\sum_{j=1}^{k}p_j^{(i)}\right|\right\} \tag{9}$$

$$overlap = \frac{|R^g \cap R^r|\|Z_0 - Z|}{R^g \cup R^r} \tag{10}$$

$$P(D_i, w_j) = P(d_i)P(w_j|d_i);$$
$$P(w_j|d_i) = \sum_{k=1}^{K}P(w_j|z_k)P(z_k|d_i) \tag{11}$$

where, P is the position of G-BS in the cluster; Z_0, Z are the locations of service U-BS and interference U-BS in the air cluster, respectively; R is a typical ground user (GUE); K is the distance from the horizontal projection of the air service U-BS of the typical GUE to the representative cluster center and the typical GUE; D is the distance from the horizontal projection of the air interference U-BS to the typical GUE and the representative cluster center; G-BSx is the position of the G-BS between clusters; v = x is the distance from the typical GUE to the G-BSx of the cluster center; V = x is the distance from the plane projection of the air interference U-BS between clusters to the typical GUE and the cluster center G-BSx. The heterogenous network of G-BS and U-BSs can be described as:

$$U = \left\{P_1|D, L, f_2, Q, d, l \quad P_2|f_1, \mu \quad P_3|N, M, I\right\} \tag{12}$$

where, D is the constraint of the feature change of the objective function. From the constraint, it is possible to derive the feature error of clear and occluded samples, and reduce the error through optimization. When the error is reduced to a certain extent, the feature of the clear sample can be approximated to the mean feature of the occluded sample. Furthermore, we have:

$$L(Y_i, y_i) = -\frac{1}{n}\sum_{i=1}^{n}[Y_i\log(y_i) + (1 - Y_i)\log(1 - y_i)] \tag{13}$$

$$H(x) = F(x) + x \tag{14}$$

where, H is the transmitting antenna at G-BS; t = u is the transmitting antenna at U-BS; f is the receiving antenna at GUE. The maximum transmission gain received by a typical GUE from the layer I transmitter can be described by:

$$\Delta_{ikjl} = \sum_{\delta=1}^{n}\Delta_{ikjl}(\varepsilon) \tag{15}$$

According to the total transmission gain that could be received by a typical GUE from the layer I transmitter T, the objective function can be obtained by:

$$\Delta_{ikjl}(\varepsilon) = \begin{cases} 0, & x_{ik}(\varepsilon) = N/A \quad or \quad x_{jl}(\varepsilon) = N/A; \\ 1, & x_{ik}(\varepsilon) = N/A \quad and \quad x_{jl}(\varepsilon) = N/A \end{cases} \tag{16}$$

$$x_i = \sum_{j=1}^{n}\omega_{ij}y_j - \theta_i \tag{17}$$

The advantages of the multi-layer network can be highlighted through cascading. Therefore, two cascading schemes, namely, two-layer cascading scheme, and our four-layer cascading scheme, were compared with each other. The former only considers the typical GUE cascading with GBS/UBS within the cluster, while latter fully considers all cases of typical GUE cascading with GBS/UBS within and between clusters. In other words, the 4-layer cascading scheme helps to quantify the impact of GBS/UBS on the performance of UAV-assisted multi-layer millimeter wave heterogeneous network. In addition, both schemes follow the maximum binary recursive portioning (BRP) criterion, that is, the typical GUE is cascaded with the G-BS/U-BS which provides the strongest mean BRP in the long term. Therefore, when a typical GUE is cascaded with the nearest G-BS/U-BS at x in layer I, the following problem will occur:

$$P(d_i, w_j) = P(d_i) P(w_j | d_i);$$
$$P(w_j | d_i) = \sum_{k=1}^{K} P(w_j | z_k) P(z_k | d_i) \tag{18}$$

where, $P(d_i, w_j)$ is the minimum path loss from a typical GUE to layer K. According to the cascading standard, the cascading probability refers to the probability of typical GUE cascading with G-BS/U-BS in LOS/NLOS state in layer I. Since the G-BS/U-BS of layers 0 and 2 are in the representative cluster, and only one G-BS/U-BS exists in the cluster nearest to the typical GUE, the link state between the typical GUE and the G-BS/U-BS in the cluster can generally be changed by adjusting the position and height of the UAV. Therefore, the link state is either LOS or NLOS. Treating the path losses of layers 0 and 2 as two states, the probability of the typical GUE cascading with G-BS/U-BS can be defined as:

$$UBS = f(W^e D_1 + \delta^e) \tag{19}$$

Taking U-BS as the input, the feature constraint can be added through the objective function to make the occluded sample close to the eigenvector of the clear sample. Then, the deviation between the target value and the actual output, and the deviation and characteristic error between the calculated target value and the actual output, were used as the characteristic parameters. In addition, the forward propagation of the network was completed through the primary and secondary approximation of the front and back terms. After the training, the weights and thresholds were determined, completing network modeling.

3. DESIGN OF HETEROGENEOUS NETWORK FOR MULTI-SOURCE DATA FUSION AND TARGET TRACKING

3.1 Modeling philosophy

Based on millimeter wave heterogeneous network, this paper constructs a multi-source data fusion and target tracking model. The core of the model is DM-DQN algorithm. Drawing on the heterogenous network of CNN, the length of input vector (time window) was expanded from 25 to 31 through image filling, in order to ensure the length matching between input and output vectors, and retain the temporal features of

eye tracking data as much as possible. The purpose is to increase the recognition range. In addition, 5G network technology was adopted to improve the recognition efficiency and fluency of multi-source data fusion and target tracking of the system. Further, the network density and link state of the system were tested, and the stability and iterative relationship of DM-DQN were analyzed in details.

3.2 Model construction

In the multi-source data fusion and target tracking model for heterogenous network, an attention module was added to each branch of the CNN. In this way, the learning focuses on a specific distinguishable local area, rather than evenly distributed on the overall view representation. Admittedly, the CNN feature output contains fine-grained information, which can be distinguished based on fine-grained details. However, there is a huge feature noise arising from the feature misalignment between the two branches, and the limited semantic perception information of each fine-grained feature. The energy functions based on Euclidean distance typically include comparison loss or triple ranking loss. The functions depend on the calculated distance per element, and are very sensitive to dislocation. The use of these two loss functions requires the eigenvectors to be completely aligned according to the elements. But this requirement goes against the reality. To solve this problem, this paper proposes a loss metric based on higher-order learnable energy function (HOLEF). The loss is derived from a pair of input vectors and the triple state loss formed by weighted external subtraction between vectors. This energy function can compute the external subtraction between two eigenvectors by comparing sketches and photos, and thus measure the feature difference between two domains, element by element. With this loss function, the retrieval accuracy could be improved by doubling the retrieval time, i.e., the efficiency would be greatly reduced.

Owing to the limited training data and the overemphasis on the loss of discrimination, it is difficult to capture all domain-invariant information by learning only two domains embedded in the common space model. These domains cannot be effectively extended to test domains different from training data. This brings the disparity and inaccuracy between training domain and test domain. Therefore, this paper proposes a new hybrid discriminant generative model, forcing the learned embedding space to retain domain-invariant information, which is useful for cross-domain reconstruction. To verify the effectiveness of the cross-modal retrieval method in SBIR, the probability method was employed to model the joint multi-modal data distribution and learn the multi-modal correlation; subspace learning is chosen to construct the common subspace, and map the multi-modal data to this subspace for cross modal matching. Image retrieval is a cross-pattern matching problem that can be solved by learning a joint embedding space, which can retain the shared semantic content between photos and image patterns. A CVAE was adopted to decompose each image into a part of shared semantic content and a part of unique style of the image painter. The CVAE was meta-trained by two additional style adaptive components, such that our model could dynamically adapt to any invisible user style: a set of feature transformation layers was added to the encoder, and a regulator was added to separate the potential encoding of semantic content. Experiments show that subspace learning can effectively model the gap between sketch and photo.

4. EXPERIMENTS AND RESULTS ANALYSIS

4.1 Energy efficiency of multi-target heterogenous network

As shown in Figure 1, the energy efficiency of the network changes with target density, at the user rate demand of 0.1, 0.5, or 1. When the user rate demand remained constant, network interference and energy consumption increased with target density, causing a gradual decline in network energy efficiency. When the target density remained constant, the network needed more transmission power to meet the rising user rate demand, which also leads to a downward trend of network energy efficiency. Therefore, the proposed DM-DQN can dynamically adjust the network state according to user QoS and optimize the energy efficiency of the network.
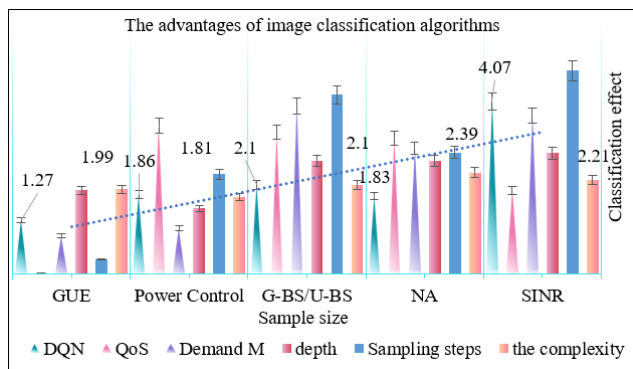

Figure 1. Advantages of image classification algorithms

As shown in Table 1, the higher the target density, the greater its influence on network performance. In DM-DQN, agents constantly interact with the environment. The RB allocation strategy and the corresponding power allocation strategy are taken as network actions to optimize the network performance, and their interaction is considered in a comprehensive manner. Through continuous trial and exploration, agents gradually approach the best strategy. After DNN training, the agents can adaptively adjust the strategy for data fusion and target tracking of the heterogenous network, according to the changing environment of the network. That is

why our network performs better than data mining algorithm and two-stage algorithm.

As shown in Figure 2, the total energy efficiency of the network changed with target density, under the user rate demand of 0.5m. The growing density of nano targets in the network suppressed the overall network energy efficiency of every algorithm. The reason is that more nano targets push up network interference and energy consumption, resulting in poorer network performance.

As shown in Table 2, DM-DQN was more energy efficient than the typical data mining algorithm and two-stage algorithm, and achieved comparable effect as the efficiency traversal algorithm with the optimal energy efficiency. This is because the two-stage algorithm optimizes RB allocation and power control in two separate stages. The RB allocation stage could avoid some network interference, but the allocation strategy is predetermined during the power control. Thus, the overall performance of this algorithm can only be improved by a limited margin.
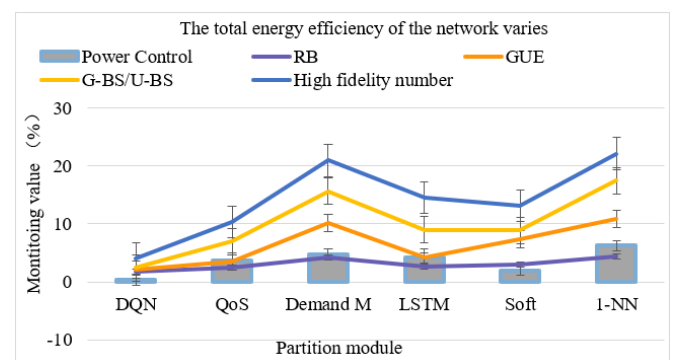

Figure 2. Variation of total energy efficiency of the network

4.2 Influence of link cascading on target tracking

As shown in Figure 3, the probability of a typical GUE and G-BS/U-BS in LOS link state was generally higher than that in NLOS link state. At a small Rb value, the probability of typical GUE cascading with G-BS/U-BS within the representative cluster was usually greater than that of typical GUE cascading with G-BS/U-BS between clusters.
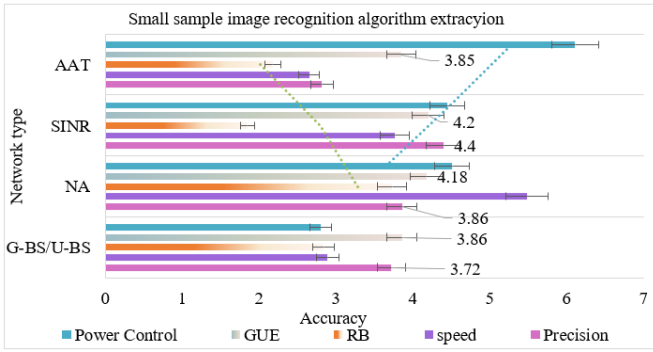
Table 1. Continuous interaction between agents and environment

Item	DM-DQN	QoS	Demand m	Depth	Sampling steps	Complexity
GUE	1.27	0.01	0.89	1.97	0.34	1.99
Power Control	1.86	3.49	1.07	1.54	2.35	1.81
G-BS/U-BS	2.1	3.33	3.96	2.67	4.22	2.1
NA	1.83	3.2	2.96	2.67	2.86	2.39
SINR	4.07	1.97	3.73	2.85	4.79	2.21
AAT	1.46	1.17	6.53	2.2	3.28	6.95

Table 2. Superiority of DM-DQN in energy efficiency

Algorithm	RB	GUE	Power Control	G-BS/U-BS	High fidelity number
DM-DQN	1.74	0.45	0.31	0.18	1.7
QoS	2.39	1.12	3.74	3.46	3.42
Demand m	4.29	5.89	4.75	5.44	5.34
LSTM	2.63	1.64	4.19	4.75	5.53
Soft	2.95	4.35	2	1.61	4.27
1-NN	4.46	6.48	6.29	6.53	4.66

Figure 3. Performance of small sample image recognition

Table 3. Relationship between cascading probability and G-BS density

Algorithm	Precision	Speed	RB	GUE	Power control
G-BS/U-BS	3.72	2.89	2.84	3.86	2.8
NA	3.86	5.48	3.73	4.18	4.51
SINR	4.4	3.77	1.85	4.2	4.45
AAT	2.82	2.65	2.18	3.85	6.11

As shown in Table 3, the cascading probability A1 and A3 of typical GUE and G-BS/U-BS between clusters increased with G-BS density, while the cascading probability NA of typical GUE and G-BS/U-BS within representative clusters decreased with the increase of G-BS density.

Figure 4. Model of image target detection

As shown in Figure 4, under the user QoS constraint in DM-DQN, the higher the target density, the greater the network interference, and the more transmission power is needed to ensure user rate. Therefore, the lower the target density, the smaller the gap between DM-DQN and the enumeration algorithm. When the user rate demand was 0.5m, the total user delay in the image increased with target density.

Table 4. Effectiveness of DM-DQN on ensuring user QoS

Pixel size	Uplink density	Downlink density	Character density	Pixel density
480	3.5	3.41	4.11	4.06
720	3.93	4.35	1.02	1.4
1,080	3.95	1.85	2.82	3.87
1,440	6.71	1.43	1.61	5.77

As shown in Table 4, after 160s of tracking, the modified mean error of the data mining based on the proposed model was only 1.5m, a sign of ideal tracking effect. Therefore, DM-DQN performs better than other algorithms in total user delay. DM-DQN can improve energy efficiency of the network, and guarantee the user QoS. With the increase of target density, the number of users in the network increased, the network interference intensified, and the total user delay gradually

lengthened. As for the enumeration algorithm, the optimal energy efficiency is taken as the optimization objective to maximize the energy efficiency of the whole system. When the target density increased, the rate of individual users declined, and the total network delay grew, resulting in a longer delay of the enumeration algorithm. The proposed DM-DQN effectively reduces the network interference and ensures the user rate by taking the total user delay as a part of the return function, and optimizing the RB and power jointly, with the RB allocation and power allocation strategies as executive actions.

Table 5. Iterative convergence of algorithms

Item	Number	GUE	Power control	G-BS/U-BS	NA	SINR
Precision	10	1.89	1.8	1.88	1.02	1.12
Speed	25	1.52	3.48	2.24	1.89	1.78
Range	50	2.03	2.3	2.53	5.11	4.27
Complexity	80	3.76	4.39	3.11	4.26	5.39
Performance	100	2.69	4.16	1.91	2.23	2.21
Accuracy	120	4.04	1.89	2.15	6.34	5.29

As shown in Table 5, DM-DQN gradually converged after nearly 100 iterations. In the first 50 iterations, DM-DQN was outperformed by structured data mining. In the early phase, structured data mining learns from the initial feedback, while DM-DQN only randomly selects actions and stores the feedback in the playback experience pool. After 50 iterations, DM-DQN began to learn from the stored feedback. After 100 iterations, both DM-DQN and structured data mining tended to be stable, with the former having an edge in performance. Overall, DM-DQN converged faster and performed better than the typical structured data mining algorithm.
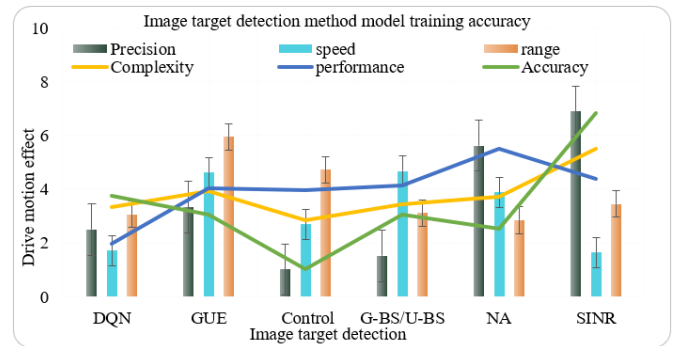

Figure 5. Joint RB allocation and power control

The optimization problem of joint RB allocation and power control was proposed for the ultra-dense heterogeneous network, in the light of user QoS, aiming to reduce the same layer and cross layer interference and improve the energy efficiency of the network. To avoid the high complexity of traditional algorithms, the DM-DQN framework was introduced, and a reward function was defined to optimize network energy efficiency and ensure user QoS. As shown in Figure 5, our DM-DQN algorithm guaranteed the QoS of users, and achieved closer-to-optimal performance than typical data mining, two-stage algorithm, and enumeration algorithm. Next, multi-agent-based distributed resource management would be studied, trying to reduce network interference and improve energy efficiency through multi-agent cooperation.

As shown in Figure 6, 5G/B5G network architecture could be improved from fixed ground infrastructure to air mobile

connection. Millimeter wave communication is the preferred way to realize 5G/B5G ultra reliable and low delay communication. In the UAV-assisted wireless communication system, UAVs can fly out of the blocking area to establish LOS links, which can overcome the penetration loss and facilitate the transmission of millimeter wave signals. Hence, a multi-layer heterogeneous network was formed with the coexistence of ground cellular network and UAV network. Due to the irregularity of the location for ground target (G-BS)/UAV target (U-BS), it is important to realize accurate modeling and simplified analysis through mathematical methods, such as random geometry and spatial point process. In addition, the spatial coupling capacity is insufficient between G-BS/U-BS and UE.

Figure 6. Air mobile connection of 5G/B5G network architecture
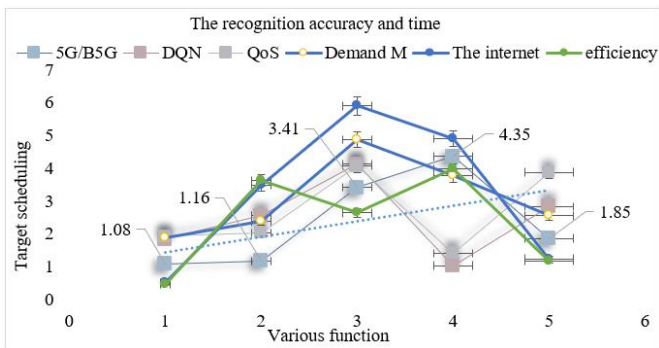
Figure 7. Recognition accuracy and time

As shown in Figure 7, with the increase of SINR threshold, the AAT could reach the maximum, and the optimal SINR threshold stood at 0. When the SINR threshold was above zero, the AAT decreased with the rise of the threshold. From the AAT corresponding to the SINR threshold of 0 in Figure 7, it can be found that the available AAT decreased with the growing m value.

Table 6. Downlink available to G-BS/U-BS in 4-layer cascade scheme

Item	AAT	DM-DQN	QoS	Demand m	SINR	Efficiency
UAV	1.08	1.83	1.92	1.88	0.51	0.45
EGNN	1.16	2.55	2.02	2.36	3.48	3.62
SFA	3.41	4.11	4.06	4.86	5.89	2.63
DML	4.35	1.02	1.4	3.76	4.9	3.98
RNN	1.85	2.82	3.87	2.54	1.23	1.16

As shown in Table 6, when the SINR threshold remained the same, the downlink AAT that can be obtained by G-BS/U-BS in the 4-layer cascade scheme was greater than that of the 2-layer cascade scheme. As a result, the total AAT that can be obtained by the 4-layer cascade scheme was higher than that of the 2-layer cascade scheme. This confirms that the proposed

4-layer cascade scheme helps to increase the available AAT and optimize the performance of the UAV-assisted multi-layer millimeter wave heterogeneous network.
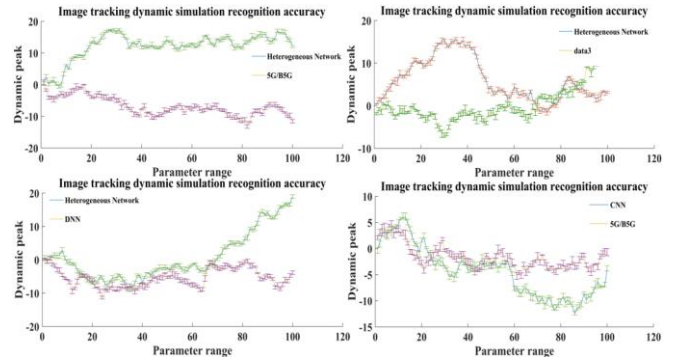

Figure 8. Dynamic recognition accuracy of image tracking

Figure 8 compares the basic network, model structure, usage method, loss function, test dataset, retrieval accuracy, and pixel loss (which ensures that the pixel information is preserved in the acquired embedding) of four representative generalization methods of depth target tracking classes. Ideally, the sketch domain and the image domain can be fully aligned in the common pixel space. Under the ideal situation, better target tracking results can be obtained by properly retaining more effective pixels in the learning process, thereby completing the retrieval of detail pixels.
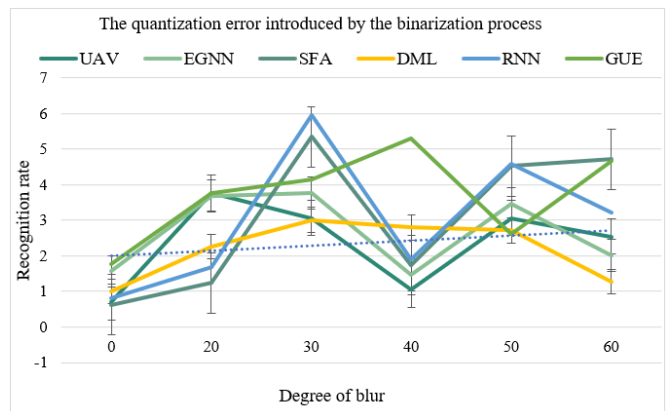

Figure 9. Quantization error introduced through binarization

As shown in Figure 9, the quantization error introduced by the binarization of hash algorithm destroyed the domain-invariant information and the pixel consistency across domains. Therefore, this paper also proposes a generating hash algorithm to reconstruct the zero-learning knowledge representation. However, a high retrieval cost was incurred by the inefficiency of Kronecker fusion layer. It is often necessary to use expensive sketch image pairs, e.g., sketchy dataset, to better align the pixels in the common mapping space between the sketch domain and the image domain.

As shown in Figure 10, a typical GUE and G-BS/U-BS was much more likely to be in LOS link state than to be in NLOS link state, that is, LA was usually greater than 0. In addition, the density G-BS did not greatly affect the cascading probability, especially when the typical GUE and the G-BS/U-BS in the representative cluster were in the LOS link state. In this case, the cascading probability was basically invariant with the rising G-BS density. When the mean number of

cluster members m was different, the authors discussed the relationship between AAT and SINR threshold, and compared the AATs between different cascading schemes. The discussion shows that the rise of SINR threshold does not always benefits AAT improvement. When the SINR threshold is relatively small, the AAT increases with SINR threshold.
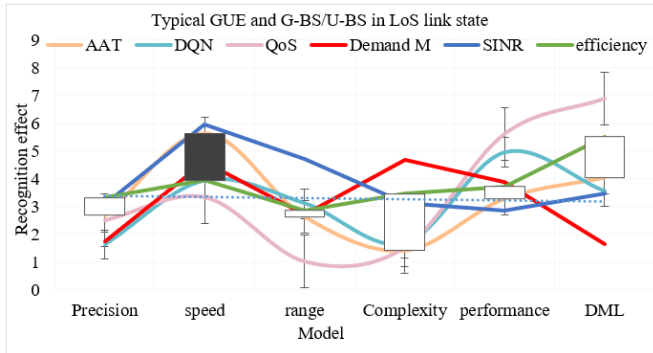

Figure 10. Typical GUE and G-BS/U-BS in LOS link state

5. CONCLUSIONS

In data mining, classification loss, cyclic consistency loss, and counter loss are often integrated to keep the cyclic consistency of each branch. In this way, the model only needs to learn under class level supervision, eliminating the need for expensive sketch image pairs or memory fusion layer. This paper combines domain separation loss with the gradient inversion layer (GRL). The former bridges the gap between domains by forcing the network to learn domain independent embedding. The latter encourages the encoder to extract mutual information from sketches and photos.

With the increase of target density, the number of users in the network increases, the network interference intensifies, and the total user delay gradually lengthens. In the enumeration algorithm, the optimal energy efficiency is taken as the optimization objective to maximize the energy efficiency of the whole system. When the target density increases, the rate of individual users decreases, and the total delay of the whole network rises, resulting a greater delay. Taking the total user delay as a part of the return function, the proposed DM-DQN optimizes the RB and power jointly by taking the RB allocation and power allocation strategies as executive actions, which effectively reduce the network interference and ensure the user rate. Our algorithm can improve the network energy efficiency and guarantee the QoS of users.

To adapt to the various scenarios of 5G/B5G, wireless communication technology is developing rapidly. As a common 5G/B5G scenario, hotspot often occurs in the area with a high concentration of mobile users, which causes data rate to surge in a short time. To satisfy this demand, 5G/B5G network must be transformed from coverage-centered macro cell deployment to capacity-centered small cell deployment. The latter is generally user-oriented, and composed of various low-power targets (BSs). The performance of 5G/B5G can be significantly improved by fusing multi-target tracking with millimeter wave. Currently, UAVs provide a new communication alternative in civil and military fields. Compared with the traditional ground cellular communication, the highly mobile UAVs can be quickly deployed to establish communication in the hotspot, forming an air-to-ground

channel with a high probability of setting up LOS links. Therefore, the future research will develop a power allocation model to improve the target capacity, analyze the model features, and design the corresponding fast algorithm for the 3D multi-target tracking problem of networked radar in clutter background.

REFERENCES

[1] Yang, X.J. (2012). Review of distributed decision fusion in wire-less sensor networks. Computer Engineering and Applications, 48(11): 1-6.

[2] Chen, H., Han, C.Z. (2016). Sensor control strategy for maneuvering multi-target tracking. Acta Automatica Sinica, 42(4): 512-523. http://dx.doi.org/10.16383/j.aas.2016.c150529

[3] Sun, S., Lin, H., Ma, J., Li, X. (2017). Multi-sensor distributed fusion estimation with applications in networked systems: A review paper. Information Fusion, 38: 122-134. https://doi.org/10.1016/j.inffus.2017.03.006

[4] Vo, B.N., Ma, W.K. (2006). The Gaussian mixture probability hypothesis density filter. IEEE Transactions on Signal Processing, 54(11): 4091-4104. https://doi.org/10.1109/TSP.2006.881190

[5] Vo, B.T., Vo, B.N., Cantoni, A. (2007). Analytic implementations of the cardinalized probability hypothesis density filter. IEEE Transactions on Signal Processing, 55(7): 3553-3567. https://doi.org/10.1109/TSP.2007.894241

[6] Li, T., Wang, X., Liang, Y., Pan, Q. (2020). On arithmetic average fusion and its application for distributed multi-Bernoulli multitarget tracking. IEEE Transactions on Signal Processing, 68: 2883-2896. https://doi.org/10.1109/TSP.2020.2985643

[7] Chen, H., He, Z.L., Deng, D.M., Li, G.C. (2020). Sensor control using Cauchy-schwarz divergence via Gaussian mixture multi-Bernoulli filter. Acta Electonica Sinica, 48(4): 706-716. https://doi.org/10.3969/j.issn.0372-2112.2020.04.012

[8] Ren, W., Beard, R.W., Atkins, E.M. (2007). Information consensus in multivehicle cooperative control. IEEE Control Systems Magazine, 27(2): 71-82.

https://doi.org/10.1109/MCS.2007.338264

[9] Li, T., Fan, H., García, J., Corchado, J.M. (2019). Second-order statistics analysis and comparison between arithmetic and geometric average fusion: Application to multi-sensor target tracking. Information Fusion, 51: 233-243. https://doi.org/10.1016/j.inffus.2019.02.009

[10] Li, G., Battistelli, G., Yi, W., Kong, L. (2020). Distributed multi-sensor multi-view fusion based on generalized covariance intersection. Signal Processing, 166: 107246. https://doi.org/10.1016/j.sigpro.2019.107246

[11] Wang, B., Yi, W., Hoseinnezhad, R., Li, S., Kong, L., Yang, X. (2016). Distributed fusion with multi-Bernoulli filter based on generalized covariance intersection. IEEE Transactions on Signal Processing, 65(1): 242-255. https://doi.org/10.1109/TSP.2016.2617825

[12] Wang, B.L., Yi, W., Li, S.Q., Kong, L.J., Yang, X.B. (2018). Consensus for distributed multi-Bernoulli filter. Signal Processing, 34(1): 1-12. https://doi.org/10.16798/j.issn.1003-0530.2018.01.001

[13] Li, T., Corchado, J.M., Sun, S. (2018). Partial consensus and conservative fusion of Gaussian mixtures for distributed PHD fusion. IEEE Transactions on Aerospace and Electronic Systems, 55(5): 2150-2163. https://doi.org/10.1109/TAES.2018.2882960

[14] Li, T., Liu, Z., Pan, Q. (2019). Distributed Bernoulli filtering for target detection and tracking based on arithmetic average fusion. IEEE Signal Processing Letters, 26(12): 1812-1816. https://doi.org/10.1109/LSP.2019.2950588

[15] Gao, L., Battistelli, G., Chisci, L. (2020). Multiobject fusion with minimum information loss. IEEE Signal Processing Letters, 27: 201-205. https://doi.org/10.1109/LSP.2019.2963817

[16] Lu, J.H., Han, X., Li, J.X. (2016). Consensus-based distributed fusion estimator with communication bandwidth constraints. Control and Decision, 31(12): 2155-2162. https://doi.org/10.13195/j.kzyjc.2015.1481

[17] Kamal, A.T., Farrell, J.A., Roy-Chowdhury, A.K. (2013). Information weighted consensus filters and their application in distributed camera networks. IEEE Transactions on Automatic Control, 58(12): 3112-3125. https://doi.org/10.1109/TAC.2013.2277621

[18] Ilić, N., Stanković, M.S., Stanković, S.S. (2013). Adaptive consensus-based distributed target tracking in sensor networks with limited sensing range. IEEE Transactions on Control Systems Technology, 22(2): 778-785. https://doi.org/10.1109/TCST.2013.2256787

[19] Li, T., Elvira, V., Fan, H., Corchado, J.M. (2018). Local-diffusion-based distributed SMC-PHD filtering using sensors with limited sensing range. IEEE Sensors Journal, 19(4): 1580-1589. https://doi.org/10.1109/JSEN.2018.2882084

[20] Vo, B.N., Singh, S., Doucet, A. (2005). Sequential Monte Carlo methods for multitarget filtering with random finite sets. IEEE Transactions on Aerospace and electronic systems, 41(4): 1224-1245. https://doi.org/10.1109/TAES.2005.1561884

[21] Mahler, R.P. (2004). "Statistics 101" for multisensor, multitarget data fusion. IEEE Aerospace and Electronic Systems Magazine, 19(1): 53-64. https://doi.org/10.1109/MAES.2004.1263231

[22] Li, T.C., Fan, H.Q., Sun, S.D. (2015). Particle filtering: Theory, approach, and application for multitarget tracking. Acta Automatica Sinica, 41(12): 1981-2002. https://doi.org/10.16383/j.aas.2015.c150426

[23] Ye, J.C., Bresler, Y., Moulin, P. (2000). Asymptotic global confidence regions in parametric shape estimation problems. IEEE Transactions on Information Theory, 46(5): 1881-1895. https://doi.org/10.1109/18.857798

[24] Liu, G., Chen, X. (2013). Optimal sub pattern assignment probability metric for multi-target tracking algorithm. Computer Engineering, 39(5): 293-296.

[25] Sun, L., Yu, H., Fu, Z., He, Z., Tao, F. (2020). Fast measurement partitioning algorithm for multiple extended target tracking. Electronics Letters, 56(16): 832-835. https://doi.org/10.1049/el.2020.0984