

Article

Identification of Large Yellow Croaker under Variable Conditions Based on the Cycle Generative Adversarial Network and Transfer Learning

Shijing Liu ^{1,2,*}, Cheng Qian ¹, Xueying Tu ¹, Haojun Zheng ³, Lin Zhu ^{1,*}, Huang Liu ¹  and Jun Chen ¹

¹ Fishery Machinery and Instrument Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200092, China; qiancheng@fmiri.ac.cn (C.Q.); tuxueying@fmiri.ac.cn (X.T.); liuhuang@fmiri.ac.cn (H.L.); chenjun@fmiri.ac.cn (J.C.)

² Sanya Oceanographic Institution, Ocean University of China, Sanya 572011, China

³ School of Navigation and Naval Architecture, Dalian Ocean University, Dalian 116023, China; zhenghaojun2023@163.com

* Correspondence: liushijing@fmiri.ac.cn (S.L.); zhulin@fmiri.ac.cn (L.Z.); Tel.: +86-156-1867-6160 (S.L.)

Abstract: Variable-condition fish recognition is a type of cross-scene and cross-camera fish re-identification (re-ID) technology. Due to the difference in the domain distribution of fish images collected under different culture conditions, the available training data cannot be effectively used for the new identification method. To solve these problems, we proposed a new method for identifying large yellow croaker based on the CycleGAN (cycle generative adversarial network) and transfer learning. This method constructs source sample sets and target sample sets by acquiring large yellow croaker images in controllable scenes and actual farming conditions, respectively. The CycleGAN was used as the basic framework for image transformation from the source domain to the target domain to realize data amplification in the target domain. In particular, IDF (identity foreground loss) was used to optimize identity loss judgment criteria, and MMD (maximum mean discrepancy) was used to narrow the distribution between the source domain and target domain. Finally, transfer learning was carried out with the expanded samples to realize the identification of large yellow croaker under varying conditions. The experimental results showed that the proposed method achieved good identification results in both the controlled scene and the actual culture scene, with an average recognition accuracy of 96.9% and 94%, respectively. These provide effective technical support for the next steps in fish behavior tracking and phenotype measurement.

Keywords: fish recognition; CycleGAN; foreground mask loss; maximum mean discrepancy; transfer learning



Citation: Liu, S.; Qian, C.; Tu, X.; Zheng, H.; Zhu, L.; Liu, H.; Chen, J. Identification of Large Yellow Croaker under Variable Conditions Based on the Cycle Generative Adversarial Network and Transfer Learning. *J. Mar. Sci. Eng.* **2023**, *11*, 1461. <https://doi.org/10.3390/jmse11071461>

Academic Editor: Alexei M. Orlov

Received: 27 June 2023

Revised: 20 July 2023

Accepted: 21 July 2023

Published: 22 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The large yellow croaker (*Larimichthys crocea*) is marine migratory fish of the Pacific Northwest [1]. In recent years, due to its high economic value, large yellow croaker has become one of the most commercially valuable marine fishery species in China's aquaculture production [2]. Accurate identification of large yellow croaker under variable conditions is of great significance to improve the ability of the high-throughput detection of fish phenotypes in genetic breeding and aquaculture production [3]. Affected by differences in sampling methods [4,5], illumination [6], and the farming environment [7,8], the images obtained in different farming scenarios have different domain distribution, which limits the effect of data interoperability and increases the difficulty associated with the industrial application of identification technology. In recent years, with the development of transfer learning [9,10] and person re-ID [11,12], a possible solution for the accurate identification of fish targets under variable working conditions has been provided.

With the progress of information technology such as artificial intelligence and deep learning, the identification technology of production objects, diseases, and behaviors in

the agricultural field has been continuously developed and has been widely used in different fields of the industry [13,14]. However, compared with static objects such as rice and plants [15,16], and large land-based animals such as cattle and sheep [17,18], the development of underwater freestyle moving-target recognition technology is slow, and relevant studies have mostly focused on application scenarios where specific working conditions and training data are easy to obtain [19]. To solve this problem, transfer learning technology has been introduced into the field of fish identification. For example, Zhang et al. [6] proposed a transfer learning method based on a residual network to realize unconstrained swimming fish identification. Yuan et al. [20] used a metric learning network based on a residual structure to realize 5-way, 15-shot fish target recognition, and the recognition accuracy was higher than 90%. The method based on small samples and transfer learning can effectively improve the accuracy of fish identification. However, the application of this method has certain limitations when targeting unconstrained swimming fish under obvious actual farming conditions with relatively different backgrounds and postures. This is mainly due to: (1) the variation in sampling device and scene leading to a difference in domain distribution between the target domain and source domain, which results in the ineffective use of available training data in the new recognition domain; and (2) the change in fish swimming posture leading to the dispersion of target features, and a single data source cannot cover all feature spaces, which reduces the adaptability of the algorithm to different features.

Re-identification (Re-ID) is a technology that unifies images from different source domains into the feature space of the target domain through image domain-to-domain conversion to achieve data enhancement. It is mainly used to solve the limitations of supervised methods in the application of real scenes, and has made significant progress in the field of pedestrian re-identification. For example, Wang et al. [21] used attribute features to transfer the model to an unlabeled dataset; Deng et al. [22] embedded a twin network into the CycleGAN [23] to realize image transmission from the source domain to the target domain. Ye et al. proposed RACE (robust anchor embedding) [24] and DGM (dynamic graph co-matching) [25] to solve the video-based unsupervised person re-identification problem. Tang et al. [26] used the CycleGAN and MMD methods to strengthen the retention of pedestrian identity information and narrow the distribution of the domain.

Inspired by this image domain transfer method, we proposed a large yellow croaker recognition method based on cyclic adversarial networks and transfer learning. In this study, large yellow croaker images were collected as source samples in controlled scenes in a specific environment to provide basic image samples for fish recognition in different scenes. The large yellow croaker in the scene to be identified was collected as the target domain sample. The CycleGAN was adopted as the basic model for the transfer from the source domain to the target domain. The foreground mask self-evaluation method [27] was used to optimize the evaluation effect of the model's identity loss. MMD was introduced as the loss function to improve the model's ability to adapt to the distribution of pull and push domains. Then, the expanded sample was used for transfer learning to reduce the influence of uneven distribution of transfer learning sample data on the recognition accuracy, and realize the recognition of a free-swimming large yellow croaker. Finally, the ablation test and comparison test were used to verify the effectiveness of the proposed method.

2. Proposed Method

2.1. Method Overview

In this study, the ReID method was mainly used to unify the style of fish images obtained from different scenes, increase the number of target samples to be identified in the application scene and improve the adaptability of the algorithm. Therefore, in order to preserve identity information and extract the distribution between different domains, we embedded a foreground mask loss and a MMD layer in the CycleGAN to enable image domain-to-domain transfer. In addition, transfer learning has been shown to have advantages in feature reuse, but due to the uneven distribution of pre-trained samples, the perfor-

mance of the models varies significantly in different recognition tasks. Therefore, in order to increase the recognition ability of the fish features of the transfer learning pre-trained model, we optimized the knowledge transfer process by expanding the fish dataset. The overall framework of the proposed method is shown in Table 1.

Table 1. The results of fish ablation test to source and target. We evaluated the source and target with protocol recall, specificity and mAP (%).

Method	Target to Source			Source to Target		
	Recall	Specificity	mAP	Recall	Specificity	mAP
No Transfer	52	86	56.2	24.2	90.9	57.6
Direct Transfer	47.5	100	73.8	30.3	100	65.2
CycleGAN	57.5	100	78.8	18.2	97.5	61.7
CycleGAN + IDF	60	100	80	12.1	97.5	58.9
CycleGAN + MMD	65	95	81	24.3	100	65.8
CycleGAN + IDF + MMD	77.5	100	88.7	69.5	97.5	85

As shown in Figure 1, in our algorithm, source domains (specific breeding scenarios) and target domains (ship farming scenarios) were input into the CycleGAN to generate false target domains and false source domains. In a large-scale water mass, due to the relatively sparse spatial distribution of fish, it is easy to obtain background images without fish, and common foreground extraction algorithms can segment the foreground and background more accurately. Therefore, in the conversion process, using the difference in foreground image changes to calculate identity loss, the false source domain identity information could be pulled to the target domain identity information. At the same time, the distribution of the false target domain will be pulled towards the target domain. After the translation was completed, the tagged fish source domain image was transferred to the target domain image to realize the expansion of the fish sample set in the culture scene. Finally, the transfer model was trained with the expanded data to further improve the target recognition accuracy.

2.2. CycleGAN-Based Translation

The CycleGAN is an image transformation model based on the generative adversarial network(s), which consists of two pairs of generators and discriminators. G is the mapping function from the source domain to the target domain, and $\hat{G}$ is the mapping function from the target domain to the source domain. D_S and D_T are style discriminators for the source and target domains, respectively. S and T represent the source and destination domains, respectively. The CycleGAN mainly realizes the image conversion of two different domains by minimizing the loss function, so as to realize the multi-modal conversion between domains. The objective function consists of three parts: adversarial loss, cycle consistency loss and identity loss. The purpose of the adversarial loss function is to make the generated image indistinguishable from the real image of the target domain. The adversarial loss is used to maximize the probability of the discriminator to output the image to the generator, which is used to improve the quality of the converted image and make it more realistic. Applying adversarial loss to the two mapping functions, the objectives are expressed as:

$$\mathcal{L}_{GAN}G, D_T, S, T = \mathbb{E}_{t \sim P_{data}^t} [\log D_T(t)] + \mathbb{E}_{s \sim P_{data}^s} [\log(1 - D_T(Gs))] \quad (1)$$

$$\mathcal{L}_{GAN}\hat{G}, D_S, T, S = \mathbb{E}_{s \sim P_{data}^s} [\log D_S(s)] + \mathbb{E}_{t \sim P_{data}^t} [\log(1 - D_S(\hat{G}t))] \quad (2)$$

where s and t are the source domain image and the target domain image, respectively. Since the full diversity of the target domain cannot be captured using adversarial loss alone,

the generator may produce a limited or repetitive output, and the correct mapping from a single input s to the desired output t cannot be guaranteed. Therefore, the CycleGAN uses cycle consistency loss so that the learned mapping function has periodic consistency. The cycle consistency loss improves the generator's ability to generate images that retain the original image by minimizing the difference between the original input image and the cyclic production image, thereby improving the accuracy of image conversion. The cycle consistency loss is expressed as:

$$\mathcal{L}_{cyc} G, \hat{G} = \mathbb{E}_{s \sim P_{data^s}} [\|\hat{G}(G(s)) - s\|_1] + \mathbb{E}_{t \sim P_{data^t}} [\|G(\hat{G}(t)) - t\|_1] \quad (3)$$

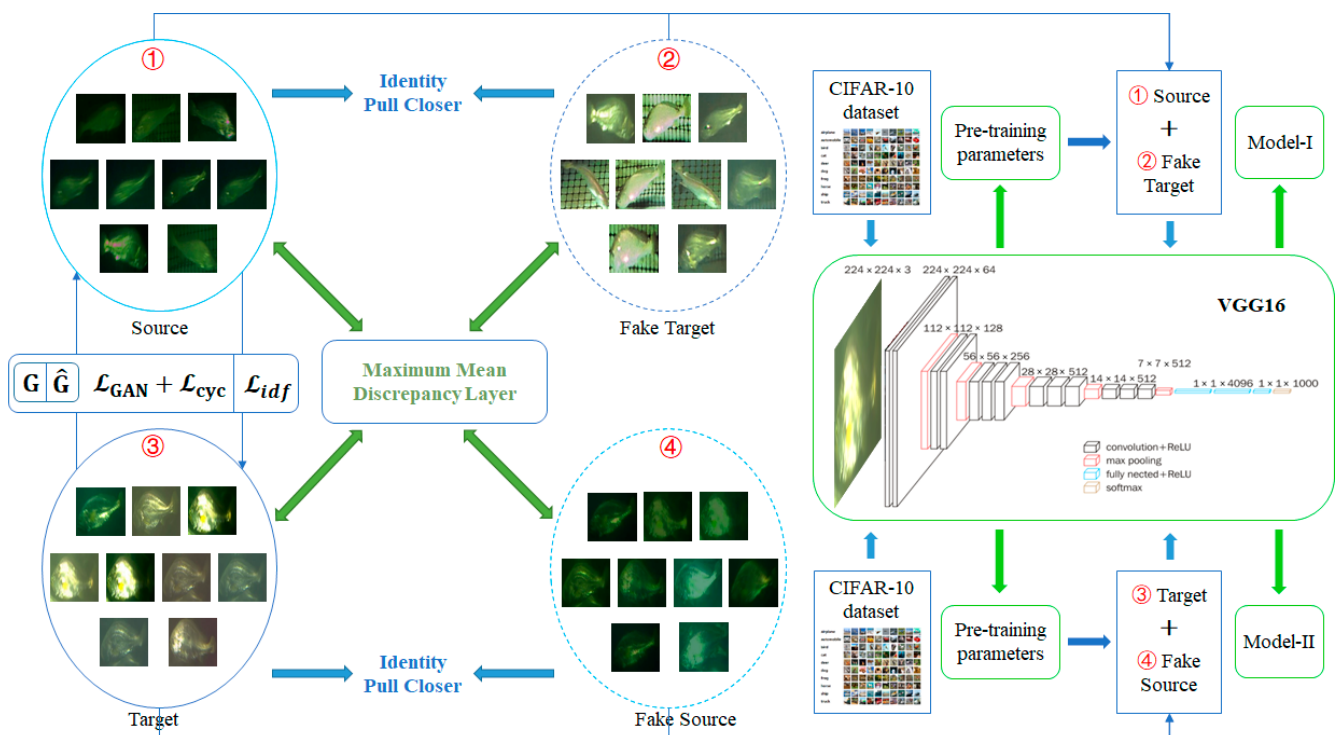


Figure 1. The framework for our method. There were two components of our method (i.e., a data transfer layer and a knowledge transfer layer). The data transfer part was mainly used to achieve sample expansion in the target domain, including CycleGAN, IDF, and maximum mean discrepancy. CycleGAN was mainly used to transfer images from the source domain to the target domain. IDF restricts CycleGAN to retaining fish identity information during the transfer process. The maximum mean discrepancy was used to narrow the distribution between the source and destination domains during transmission. Knowledge transfer was mainly used to improve the ability of the model to recognize the characteristics of fish, and the amplified data were mainly used to increase the effect of the transfer model on knowledge transfer.

2.3. Identity Foreground Loss

As part of the CycleGAN loss function, the identity loss forces the generator to not change the characteristics of the input image, but to maintain its own characteristics. The CycleGAN identity loss only uses global characteristics to count identity loss, and does not consider the impact of background noise on identity information, which leads to the mis-generation of identity during style transfer. However, under actual farming conditions, phenomena such as light absorption, scattering and diffraction caused by water turbidity reduce the feature difference between foreground fish and background noise in the image. This results in unclear identification of the fish after style conversion, which increases the risk of misidentification. To solve this problem and ensure the correct identification of fish as much as possible, we introduced the foreground constraint into the identity loss to

evaluate the changes in fish before and after migration. Due to the large volume of aquaculture water and the relatively dispersed distribution of fish under actual farming conditions, the background difference method could be easily used to obtain fish foreground images [27]. Therefore, the fish foreground images were used as the constraint conditions, and Formula (4) was used to calculate the loss of fish identity information.

$$\mathcal{L}_{IDF} = \mathbb{E}_{s \sim P_{data}s} [\| (G(s) - s) \odot M(s) \|_2] + \mathbb{E}_{t \sim P_{data}t} [\| (\hat{G}(t) - t) \odot M(t) \|_2] \quad (4)$$

where $M(s)$ and $M(t)$ represent the foreground mask of the fish image with a specific pose, and $\odot$ represents the same or logical operation.

2.4. Maximum Mean Discrepancy

For large yellow croaker images collected under different working conditions, the CycleGAN only transferred the background style of each image from the source domain to the target domain, ignoring the intra-domain distribution differences. The distribution difference provides different reference features for target recognition, which is very important in the task of target recognition with variable characteristics. Maximum mean discrepancy is mainly used to evaluate whether the distribution of two datasets is similar, and in the field of style transfer, it is mainly used to minimize the distribution difference between two networks. Therefore, the maximum mean discrepancy was used to measure the distribution difference between different sampling scenarios to solve the problem of fish sample enhancement.

$$\mathcal{L}_{MMD} = \left[\frac{1}{m(m-1)} \sum_{i \neq j}^m k(s_i, s_j) + \frac{1}{n(n-1)} \sum_{i \neq j}^n k(t_i, t_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(s_i, t_j) \right]^{\frac{1}{2}} \quad (5)$$

where k is the kernel function, m and n are the number of samples in the source and target domains, respectively, and i and j represent the coordinates of samples in specific domains. As is shown in Formula (6), the Gaussian kernel function was chosen in this paper to calculate the inner product between feature graphs.

$$k(s, s') = \exp\left(-\frac{\|s - s'\|^2}{2\sigma^2}\right) \quad (6)$$

2.5. Full Objective Function

By combining the CycleGAN, foreground mask loss and maximum mean discrepancy, we could achieve the full objective of CGAN-TM as:

$$\mathcal{L} = \mathcal{L}_{GAN}G, D_T, S, T + \mathcal{L}_{GAN}\hat{G}, D_s, T, S + \lambda_1 \mathcal{L}_{cyc}G, \hat{G} + \lambda_2 \mathcal{L}_{IDF} + \lambda_3 \mathcal{L}_{MMD} \quad (7)$$

The λ_2 and λ_3 control the weights of foreground mask loss and maximum mean discrepancy during the translation process, respectively. Detailed analysis of the parameter sensitivity is presented in Section 4.7.

2.6. Transfer Learning

According to the actual farming conditions, it is difficult to construct a sufficient field sample set according to the change in the farming environment, so the identification of large yellow croaker becomes a small-sample recognition situation. In order to simplify the complexity of the model integration application, this study adopted VGG-16 as the basic transfer learning framework, and used the CIFAR-10 dataset (open dataset, 10 categories, 60,000 images) to pre-train the model. The new training samples composed of the original small-sample data and the migrated data were used to optimize the pre-trained model parameters, and the optimized model was used to realize fish target recognition.

3. Experiments

3.1. Datasets and Evaluation Protocol

In order to evaluate the effectiveness of the method proposed in this paper, we constructed two image sample datasets: source domain and target domain. The source domain samples were collected in a recirculating aquaculture system with a controlled sampling environment, and the target domain samples were collected in an actual farming environment on an aquaculture ship. We took the source domain and the target domain as the identification scenes and verified each one.

Source area image: A total of 360 large yellow croakers with different specifications were placed in the temporary rearing tank. An underwater camera was used and the underwater depth of the camera was 40 cm. The camera was parallel to the water's surface during sampling, and the sampling was continuous for 24 h. A total of 600 images of large yellow croakers in different swimming states were selected to construct a source sample set, including 480 large yellow croaker images for training and 120 images for testing.

Target area image: We selected the "Guoxin 1" aquaculture ship, No. 1 warehouse, to collect the actual farmed fish images. The warehouse is 15 m deep and 8 m in diameter, with a total of approximately 10,000 large yellow croakers. In order to avoid the impact of fish and the influence of circulating water during sampling, a sliding rail was used to conduct continuous sampling at a depth of 4 m underwater for 1 h. A total of 300 images of large yellow croakers were obtained, of which 240 images were used for training and the remaining 60 images were used for testing.

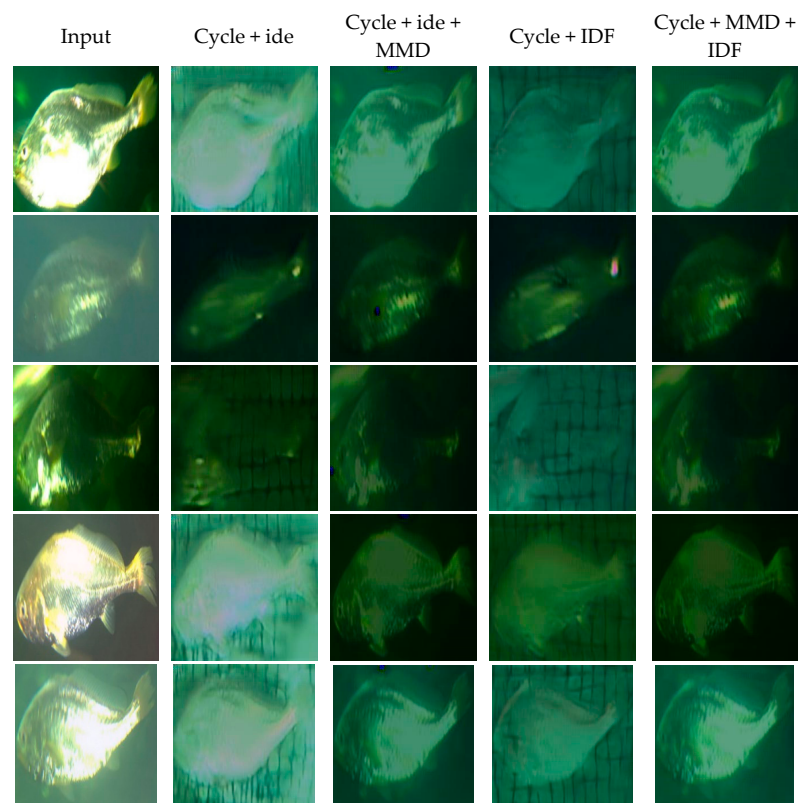
We used VGG-16 as the core framework to verify the effect of fish image transfer and the effectiveness of transfer learning in different domains. We used recall, specificity and the mean average precision (mAP) to evaluate the performance of data transfer on the source domain and target domain. Meanwhile, we selected the recall and mean average precision (mAP) to evaluate transfer learning effects.

3.2. Implementation Details

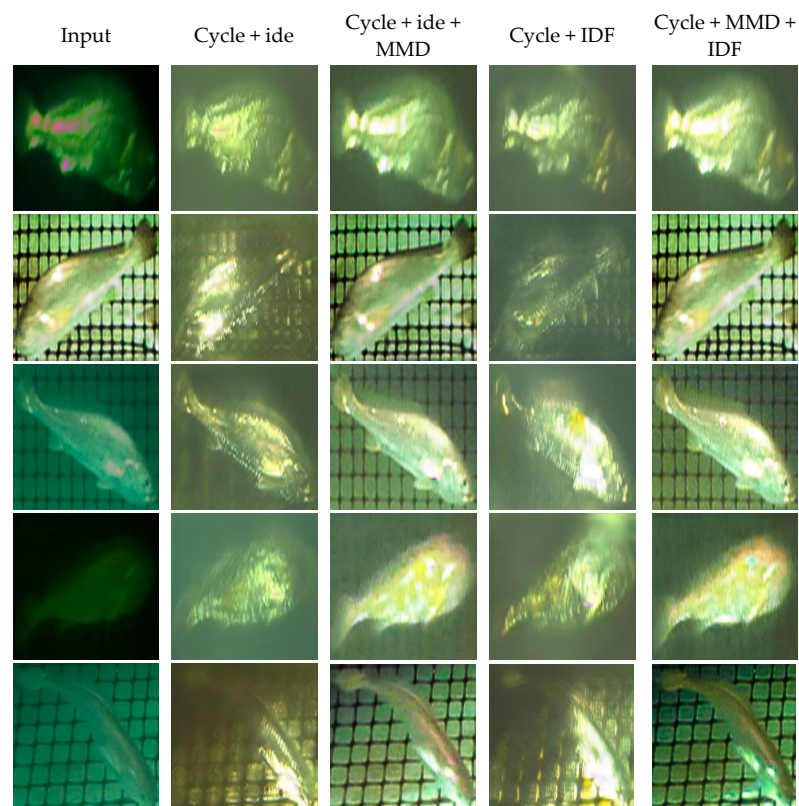
Our method was implemented using the Pytorch framework. For the CycleGAN, we used foreground mask loss instead of the identity loss function. We calculate the MMD losses using five Gaussian cores with different σ values (0.25, 0.5, 1, 2, 4) and trained them with the CycleGAN. In Equation (7), λ_1 , λ_2 and λ_3 were set as 10, 5 and (0.6, 0.8), respectively.

In order to reduce the complexity of the model framework, VGG-16, which is consistent with the CycleGAN, was selected as the transfer learning backbone network and pre-trained on the CIFAR-10 dataset. We used SGD to optimize the model, and the SGD momentum parameter set to 0.9, the weight attenuation parameter was set to 0.0005, and the learning rate was set to 0.0002. In the transfer learning stage, the original data, the generated fake data and the amplified data were used for transfer learning. Due to the small number of model parameters, freezing specific convolution layers had no obvious effect on reducing the training time, so all weight parameters were updated for transfer learning. We set the learning rate of the full connection layer to 0.01, the output dimensions to 2, the batch_size to 16, and the epoch to 60.

The GPU used was RTX A5000, the system used was Windows10, and the Pytorch version used was 1.0. Several randomly selected generated images are shown in Figure 2.



(a) Target domain to source domain



(b) Source domain to target domain

Figure 2. Source domain and target domain ablation test images. From left to right: input, CycleGAN, CycleGAN+ maximum mean discrepancy, CycleGAN+ foreground mask loss, CycleGAN+ maximum mean discrepancy + foreground mask loss (our complete method).

4. Evaluation

The main goal of fish data domain transfer was to expand the training samples, while the goal of transfer learning was to improve the fish recognition rate of a specific sample number. In order to verify the validity of this algorithm, we verified the effect of domain transfer from the source domain to the target domain, and from the target domain to the source domain.

4.1. Performance of Direct Transfer

Due to the insufficient number of samples, the model demonstrated poor performance in the source domain and the target domain. As is shown in Table 1, the recall rate of 52% and 24.24% and the mAP of 56.17% and 57.58%, in the source domain and the target domain, respectively, were achieved. However, in order to expand the number of samples, the performance of the model was slightly improved when the source domain and the target domain were directly migrated. For example, the recall rate of the data migrated to the target domain was 30.3%. Furthermore, due to the poor quality of the data in the target domain, the recall rate decreased by 4.5% after direct transfer to the source domain, and the performance decreased significantly. The main reason for this was that the source domain and target domain samples were collected under different settings, resulting in different domain distributions.

4.2. Effectiveness of the CycleGAN

As the source and target datasets are often collected in different environments, the CycleGAN is able to efficiently generate images with similar styles of datasets. Therefore, we used the CycleGAN to transfer the source domain and target domain image styles to each other, obtain fake source and fake target data. We combined the fake data with the original training data for training. As is shown in Table 1, after adding pseudo-training samples, the recall rate and mAP value of the model in the source domain increased by 10% and 5%, respectively. However, the model recall dropped to 18.2 percent in the target domain, and mAP dropped to 61.65 percent. This was mainly due to the poor quality of the target domain samples and the unsupervised transmission process of the CycleGAN, so the generated images contained a lot of noise and did not take into account the distribution of different datasets.

4.3. Necessity of Identity Foreground Loss

In order to enhance the transfer effect of fish feature information, we introduced identity foreground loss (IDF) into the CycleGAN. As is shown in Figure 2, by supervising the process of identity transfer, IDF reduces the interference of similar background features on the foreground transfer and eliminates the noise in the process of image generation. Finally, it improves the performance of the transfer model in the task of fish sample expansion. As is shown in Table 1, CycleGAN + IDF can increase the source domain recall rate to 60% and the mAP value to 80%. However, the target domain recall rate dropped to 12.1 percent and the mAP value dropped to 58.9 percent. As can be seen from Figure 2, due to the poor image quality of the target domain, the difference between the foreground and background was reduced. However, the CycleGAN + IDF was only concerned with the image difference between two different domains, but did not take into account the image difference between a specific domain, which reduced the transfer effect from the source domain to the target domain, resulting in an obvious loss of fish features in the generated images.

4.4. Importance of Maximum Mean Discrepancy

We embedded the MMD into the CycleGAN with IDF, trying to narrow the distribution by reducing the maximum mean discrepancy between the foreground in different domains. As can be seen from Table 1, the recall rate and mAP value of the model increased to 65% and 81%, respectively, after the transfer of the target domain to the source

domain. Furthermore, the increase was 24.25% and 65.75%, respectively, after the transfer of the source domain to the target domain. The results show that embedding MMD loss in the CycleGAN can successfully minimize the distribution differences between different foreground samples, which makes fish target feature extraction more efficient in different datasets. However, it can be seen from Figure 2 that the image generated by adding only MMD loss function for low-quality image data transfer still had local identity feature loss.

4.5. Practicability of Our Method

We verified the practicability of the proposed method by migrating from the source domain to the target domain and from the target domain to the source domain. Obviously, with the CycleGAN, IDF, MMD, the recall rate and mAP accuracy of the final identification results were the highest, reaching 77.5%, 88.75%, 69.5%, and 84.95%, respectively. These results increased by 30%, 15%, 39.2%, and 19.8%, respectively, compared with direct transfer. Since only 300 samples of the target set were selected, the results further prove the practicability of the proposed method in terms of its practical application.

4.6. Parameter Sensitivity

In this study, three parameters, λ_1 , λ_2 and λ_3 , control the relative importance of three target losses. We evaluated their influence on the mutual transfer between the source and target domains. λ_1 is the original parameter in the CycleGAN, and parameter 10 has been proved to be the optimal choice in the literature [23,26]. In this study, the foreground mask loss was used to optimize the identity loss function in the CycleGAN, so λ_2 could learn from the original parameters. λ_3 is a key parameter controlling MMD loss weight, so this section mainly compared the sensitivity of λ_2 and λ_3 ; the comparison results are shown in Tables 2 and 3. It is clear that both the foreground identity loss and MMD loss have been proven to be effective compared to the case of $\lambda_2 = 0$ and $\lambda_3 = 0$. From Table 3, we can see that foreground identity loss was positive when the target domain was transferred to the source domain. However, due to the poor image quality of the target domain, the features of the target to be recognized were not obvious. When the image was transferred from the source domain to the target domain, the transfer effect was poor. As can be seen from Table 4, when the weight was small, MMD loss had a significant impact on the recognition effect, and when the weight was large, the recognition effect changed slowly with the weight. Therefore, for different datasets, the values of λ_2 and λ_3 should be carefully selected due to the difference in data quality and domain distribution.

Table 2. The recall and mAP (%) results of different λ_2 values on source and target. λ_1 and λ_3 and are fixed at 10 and 0, respectively.

λ_2	Target to Source		Source to Target	
	Recall	mAP	Recall	mAP
0	57.5	78.75	18.2	61.65
2.5	58.5	79.5	12	58.8
5	60	80	12.1	58.9
7.5	58.75	78.3	11.6	58.3
10	58	70.75	11.2	58.1

Table 3. The recall and mAP (%) results of different λ_3 values on source and target. λ_1 and λ_2 and are fixed at 10 and 5, respectively.

λ_3	Target to Source		Source to Target	
	Recall	mAP	Recall	mAP
0	60	80	12.1	58.9
0.2	65.3	84	30.3	65.2

Table 3. *Cont.*

λ_3	Target to Source		Source to Target	
	Recall	mAP	Recall	mAP
0.4	70.5	85.5	54.2	77.3
0.6	77.5	88.7	60.1	80.9
0.8	77	88.4	69.5	85
1	76	88.1	69.2	84.8

Table 4. Comparison with the state-of-the-art unsupervised methods for source and target. Recall and mAP(%) were selected as the metric protocols. First results are annotated by bold type.

Methods	Target to Source		Source to Target	
	Recall	mAP	Recall	mAP
PTGAN [27]	60	80	12.1	58.9
CamStyle [28]	71.79	80.3	33.12	70.62
StarGAN [29]	82.05	88.39	24.03	66.95
Our Method	77.5	88.7	69.5	85

4.7. Comparison with State-of-the-Art Methods

We compared the proposed method with state-of-the-art methods, including inter-domain comparative transfer [27] and multi-domain joint transfer [28,29], etc. The experimental results are shown in Table 4. PTGAN (person transfer generative adversarial network) mainly considers domain differences between datasets without considering identity information loss caused by intra-domain deformation. This is similar to the method of only considering IDF loss function in the ablation experiment in this study, resulting in poor performance. CamStyle (camera style) uses label smooth regularization (LSR) to reduce the overfitting risk caused by noisy generated samples, and achieves a good effect in the target domain. However, since the loss of identity difference is not considered, the feature loss of transfer samples seriously reduces the performance of source domain recognition. StarGAN uses the mask vector to optimize the feature differences in different datasets and improve the algorithm's transfer effect among features. However, in the field of underwater free-swimming fish recognition, especially the transfer learning when the features of acquired fish images are seriously lost, the algorithm's transfer recognition effect is poor. In the target domain, recall and mAP were 24.03% and 66.95%, respectively. After the destination domain was transferred to the source domain, the mAP reached 88.39%. Compared with the above methods, this study preserved the identity information in the transmission process by introducing IDF loss, thus eliminating the background noise to a certain extent. Meanwhile, the MMD layer was adopted to learn the distribution of unlabeled datasets, thus successfully reducing the distribution difference between different foreground samples.

4.8. Effectiveness of Transfer Learning

From Table 5, we can see that the recognition accuracy of the original data was higher than that of the fake data, and the recognition accuracy of the amplified data was the highest, with recall reaching 96.5% and 87%, respectively. Overall, the recognition accuracy of fish in the source domain was higher than that in the target domain. On the whole, the fish identification accuracy in the source area was higher than that in the target area. This was mainly due to the low image quality of the target domain, resulting in more loss of identity information during data migration. By comparing the overall recognition accuracy and fish recognition accuracy, we found that, although the recall of false source domain data was low, the mAP value was high. This proves that the background recognition rate

was high. It was further demonstrated that data migration effectively distinguished between background and foreground features. On the whole, the transfer learning method effectively improved the target recognition accuracy. The recognition accuracy after the amplification of the source domain and the target domain reached 96.9% and 94%, respectively, which reflects the effectiveness of the combination of data amplification and transfer learning.

Table 5. The results of transfer learning test to source and target. We evaluated the source and target with protocol recall and mAP(%).

Training Data	Recall	mAP
Source	80.3	89.6
F (Target)	36.4	44.3
Source + F (Target)	96.5	96.9
Target	79.2	85
F (Source)	31.8	70.3
Target + F (Source)	87	94

5. Conclusions

In this paper, we proposed an improved CycleGAN and transfer learning method to recognize the large yellow croaker (*Larimichthys crocea*) in a factory ship farming scene. There are still many problems associated with a variable scene recognition task (e.g., the distribution of different datasets cannot be pulled closer during the translation process and a large number of learning samples are difficult to obtain under production conditions). To solve the first problem, we introduced the foreground ID loss and maximum mean discrepancy into the CycleGAN framework. Meanwhile, to enhance the practicality of the technology, we used transfer learning to improve recognition accuracy. We conducted extensive experiments and the results have validated the effectiveness of our method. When compared with state-of-the-art methods, the improved CycleGAN method can achieve competitive performance with a simple framework, and the final test results show that the data amplification method of domain transfer can improve the recognition accuracy of small-sample transfer learning.

Author Contributions: Conceptualization, S.L., H.L. and J.C.; data curation, H.Z.; formal analysis, C.Q. and H.Z.; funding acquisition, S.L.; methodology, S.L.; project administration, S.L. and X.T.; resources, C.Q.; software, C.Q. and H.Z.; writing—original draft, S.L.; writing—review and editing, S.L. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by Central Public-interest Scientific Institution Basal Research Fund, CAFS(NO. 2022XT06) and the earmarked fund for CARS 48.

Institutional Review Board Statement: This study complied with the regulations and guidelines established by the Animal Care and Use Committee of Fishery Machinery and Instrument Research Institute, Chinese Academy of Fishery Sciences (FMIRI-AWE-2022-001, approved on 30 September 2022).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to express their thanks to L.Z. of the Fishery Machinery and Instrument Research Institute, Chinese Academy of Fishery Sciences, for reviewing this article. The authors are thankful for the financial support received from the Central Public-interest Scientific Institution Basal Research Fund, CAFS(NO. 2022XT06) and the earmarked fund for CARS 48.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Wu, Y.; Yu, X.; Suo, N.; Bai, H.; Ke, Q.; Chen, J.; Pan, Y.; Zheng, W.; Xu, P. Thermal tolerance, safety margins and acclimation capacity assessments reveal the climate vulnerability of large yellow croaker aquaculture. *Aquaculture* **2022**, *561*, 738665. [\[CrossRef\]](#)
- Bai, Y.; Wang, J.; Zhao, J.; Ke, Q.; Qu, A.; Deng, Y.; Zeng, J.; Gong, J.; Chen, J.; Pan, Y.; et al. Genomic selection for visceral white-nodules diseases resistance in large yellow croaker. *Aquaculture* **2022**, *559*, 738421. [\[CrossRef\]](#)
- Sandford, M.; Castillo, G.; Hung, T.C. A review of fish identification methods applied on small fish. *Rev. Aquac.* **2020**, *12*, 542–554. [\[CrossRef\]](#)
- Alaba, S.Y.; Nabi, M.M.; Shah, C.; Prior, J.; Campbell, M.D.; Wallace, F.; Ball, E.B.; Moorhead, R. Class-aware fish species recognition using deep learning for an imbalanced dataset. *Sensors* **2022**, *22*, 8268. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chang, C.C.; Ubina, N.A.; Cheng, S.C.; Lan, H.Y.; Chen, K.C.; Huang, C.C. A Two-Mode Underwater Smart Sensor Object for Precision Aquaculture Based on AIoT Technology. *Sensors* **2022**, *22*, 7603. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hsiao, Y.H.; Chen, C.C.; Lin, S.I.; Lin, F.P. Real-world underwater fish recognition and identification, using sparse representation. *Ecol. Inform.* **2014**, *23*, 13–21. [\[CrossRef\]](#)
- Zhang, Z.; Du, X.; Jin, L.; Wang, S.; Wang, L.; Liu, X. Large-scale underwater fish recognition via deep adversarial learning. *Knowl. Inf. Syst.* **2022**, *64*, 353–379. [\[CrossRef\]](#)
- Liang, J.M.; Mishra, S.; Cheng, Y.L. Applying Image Recognition and Tracking Methods for Fish Physiology Detection Based on a Visual Sensor. *Sensors* **2022**, *22*, 5545. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, S.; Liu, W.; Zhu, Y.; Han, W.; Huang, Y.; Li, J. Research on fish identification in tropical waters under unconstrained environment based on transfer learning. *Earth Sci. Inform.* **2022**, *15*, 1155–1166. [\[CrossRef\]](#)
- Xu, X.; Li, W.; Duan, Q. Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification. *Comput. Electron. Agric.* **2021**, *180*, 105878. [\[CrossRef\]](#)
- Saghafi, M.A.; Hussain, A.; Zaman, H.B.; Saad, M.H.M. Review of person re-identification techniques. *IET Comput. Vis.* **2014**, *8*, 455–474. [\[CrossRef\]](#)
- Huang, N.; Liu, J.; Miao, Y.; Zhang, Q.; Han, J. Deep learning for visible-infrared cross-modality person re-identification: A comprehensive review. *Inf. Fusion* **2022**, *91*, 396–411. [\[CrossRef\]](#)
- Shruthi, U.; Nagaveni, V.; Raghavendra, B.K. A review on machine learning classification techniques for plant disease detection. In Proceedings of the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 15–16 March 2019; pp. 281–284.
- Mahmud, M.S.; Zahid, A.; Das, A.K.; Muzammil, M.; Khan, M.U. A systematic literature review on deep learning applications for precision cattle farming. *Comput. Electron. Agric.* **2021**, *187*, 106313. [\[CrossRef\]](#)
- Duong, H.T.; Hoang, V.T. Dimensionality reduction based on feature selection for rice varieties recognition. In Proceedings of the 2019 4th International Conference on Information Technology (IncIT), Bangkok, Thailand, 24–25 October 2019; pp. 199–284.
- Chen, J.; Chen, W.; Zeb, A.; Yang, S.; Zhang, D. Lightweight inception networks for the recognition and detection of rice plant diseases. *IEEE Sens. J.* **2022**, *22*, 14628–14638. [\[CrossRef\]](#)
- Zhang, C.; Zhang, H.; Tian, F.; Zhou, Y.; Zhao, S.; Du, X. Research on sheep face recognition algorithm based on improved AlexNet model. *Neural Comput. Appl.* **2023**, *35*, 1–9. [\[CrossRef\]](#)
- Peng, Y.; Kondo, N.; Fujiura, T.; Suzuki, T.; Ouma, S.; Yoshioka, H.; Itoyama, E. Dam behavior patterns in Japanese black beef cattle prior to calving: Automated detection using LSTM-RNN. *Comput. Electron. Agric.* **2020**, *169*, 105178. [\[CrossRef\]](#)
- Barbedo, J. A Review on the Use of Computer Vision and Artificial Intelligence for Fish Recognition, Monitoring, and Management. *Fishes* **2022**, *7*, 335. [\[CrossRef\]](#)
- Yuan, P.; Song, J.; Xu, H. Fish Image Recognition Based on Residual Network and Few-shot Learning. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 282–290.
- Wang, J.; Zhu, X.; Gong, S.; Li, W. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2275–2284.
- Deng, W.; Zheng, L.; Ye, Q.; Kang, G.; Yang, Y.; Jiao, J. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 994–1003.
- Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
- Ye, M.; Lan, X.; Yuen, P.C. Robust anchor embedding for unsupervised video person re-identification in the wild. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 170–186.
- Ye, M.; Li, J.; Ma, A.J.; Zheng, L.; Yuen, P.C. Dynamic graph co-matching for unsupervised video-based person re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 2976–2990. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tang, Y.; Yang, X.; Wang, N.; Song, B.; Gao, X. CGAN-TM: A novel domain-to-domain transferring method for person re-identification. *IEEE Trans. Image Process.* **2020**, *29*, 5641–5651. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wei, L.; Zhang, S.; Gao, W.; Tian, Q. Person transfer gan to bridge domain gap for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 79–88.

28. Zhong, Z.; Zheng, L.; Zheng, Z.; Li, S.; Yang, Y. Camera style adaptation for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5157–5166.
29. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8789–8797.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.