Bridging the Gap between Laboratory and Field Experiments in American Eel Detection Using Transfer Learning and Convolutional Neural Network

Tianzhixi (Tim) Yin
Pacific Northwest National Laboratory (PNNL)
tianzhixi.yin@pnnl.gov

Xiaoqin Zang
PNNL
xiaoqin.zang@pnnl.gov

Zhangshuan (Jason) Hou
PNNL
zhangshuan.hou@pnnl.gov

Paul T. Jacobson
Electric Power Research Institute
pjacobson@epri.com

Robert P. Mueller
PNNL
robert.mueller@pnnl.gov

Zhiqun (Daniel) Deng
PNNL
zhiqun.deng@pnnl.gov

Abstract

An automatic system that utilizes data analytics and machine learning to identify adult American eel in data obtained by imaging sonars is created in this study. Wavelet transform has been applied to de-noise the ARIS sonar data and a convolutional neural network model has been built to classify eels and non-eel objects. Because of the unbalanced amounts of data in laboratory and field experiments, a transfer learning strategy is implemented to fine-tune the convolutional neural network model so that it performs well for both the laboratory and field data. The proposed system can provide important information to develop mitigation strategies for safe passage of out-migrating eels at hydroelectric facilities.

1. Introduction

1.1. Adult American eel protection

The population of American eel (*Anguilla rostrata*) has significantly declined in the last few decades. The declines have been partly attributed to adults being injured or killed by being entrained into hydropower turbines when migrating from rivers to the ocean for spawning. It is technically challenging to protect this species because of its morphological and behavioral characteristics. Currently, a common practice of dam operators is to turn off turbines and provide a relatively safe, alternate downstream passage route at night for several months during the primary downstream passage period. This reduces passage via turbines to help mitigate injuries and mortality.

Information on eel behavior and the ability to identify peak migration times and pathways can assist in optimizing hydropower dam operations and the design of eel protection technologies. This not only helps improve downstream eel passage for eel recovery but also potentially lower the costs of turning off turbines. Data analytics and machine learning techniques can be applied to develop a tool that can automatically identify, track, and enumerate untagged eels in remotely sensed data gathered from large, fast-moving rivers. Such automation could also minimize the need for laborious human review for the large data files.

1.2. Literature review

Egg *et al*. [1] compared ARIS multi-beam sonar-based and GoPro camera-based methods in detecting fish and concluded that sonar is more suitable to identify riverine fish-movement patterns than optical underwater cameras in night and turbid conditions. Moreover, acoustic imaging sonar systems have significant range capability and the ability to measure fish or other objects.

In 2016, Gurshin *et al*. [2] compared three sonar technologies for observing the behavior of migrating adult eels and found that the ARIS multi-beam sonar, operating with 48 beams, is the most promising among the three for identifying eels out to 16-20 meters in range. A complete description of the project and results is now freely available to the public [11]. Mueller et al. [3] identified eels in DIDSON sonar data using three machine learning classifiers and manual feature extraction. Among the three classifiers, multiple layer perceptron (feed-forward artificial neural network) performed the best. Bothmann et al. [4] conducted fish classification using sonar data obtained by DIDSON and self-defined features.

Qin *et al*. [5] constructed a convolutional neural network (CNN) incorporating principal component analysis (PCA) and a support vector machine (SVM) classifier for fish recognition in underwater camera data. Since Krizhevsky *et al*. [6] won the ImageNet competition, CNN has become the leading machine learning model for image classification. However,
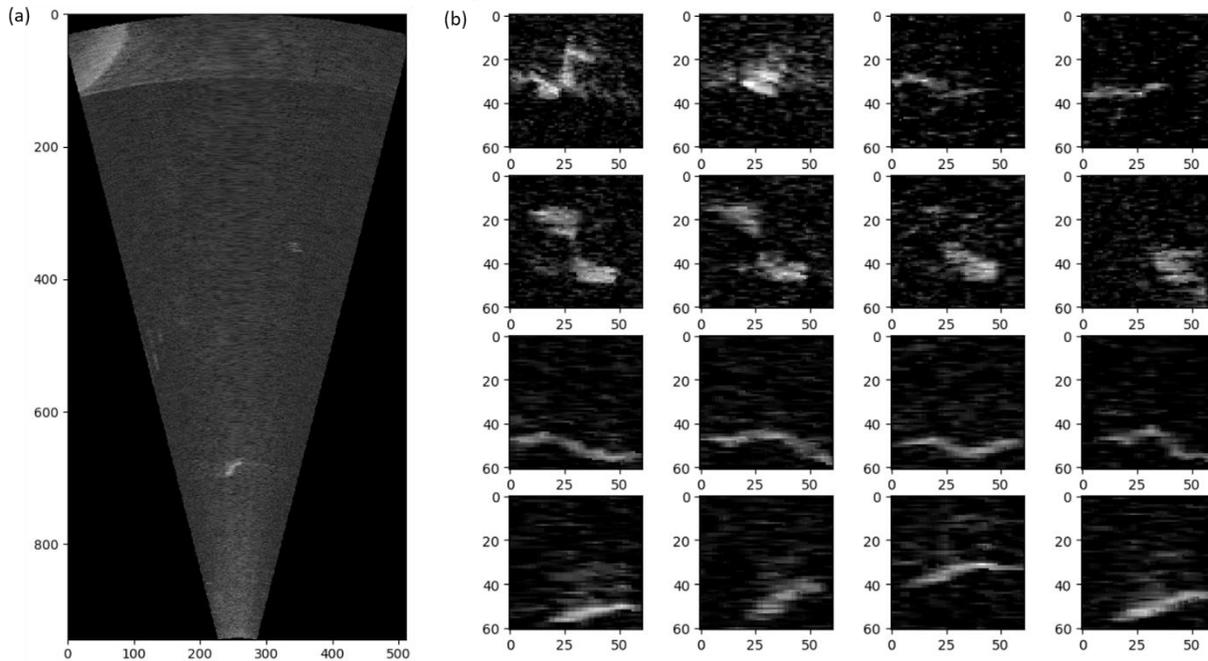
HⓎCSS

Figure 1. (a) A raw sonar image featuring an eel centered at pixel (250, 685). The structure on the upper left corner is part of the dam pier nose. (b) Example images of eels and sticks after background removal with image differencing. First row: eels in the field; second row: sticks in the field; third row: eels in the laboratory; fourth row: sticks in the laboratory.

training a deep learning model usually requires massive amount of data. For problems with limited training data, transfer learning has become a feasible option if more data from a related domain can be obtained [7].

Zhang *et al.* [8] studied synthetic aperture radar (SAR) images using CNN and transfer learning. They built a model similar to the famous visual geometry group (VGG) network [9] that classifies vehicle and transferred the model for ship classification. Sun *et al.* [10] also used transfer learning and a large-scale existing CNN model AlexNet [6] for fish classification in underwater camera images.

Developing a system-specific CNN for the automatic detection of eels in the river systems, proving that transfer learning is useful in achieving a balanced model when having limited field data and abundant lab data, and showing that the combination of sonar images and CNN is a viable option for eel monitoring are the major contributions of this study.

2. Method

2.1. Data collection

Training a robust CNN model requires sufficient representative data of the objects of detection. Due to the scarcity of migrating eels in natural environments, we construct the study with two datasets: (1) data obtained from laboratory experiments in an oval shaped water tank (7.3 m long, 3.0 m, and 2.5 m deep) in the Aquatic Research Laboratory at the Pacific Northwest National Laboratory, where four juvenile eels (yellow phase) ranging in length from 330 to 350 mm were tethered on the lower jaw so that they can swim against 0.53 and 0.76 m/s water flow within the range of the ARIS sonar (Model Explorer 1800, Sound Metrics Corp, Bellevue, WA); (2) field data collected at Iroquois Dam on the St. Lawrence River in a previous study [11], which featured adult eels ranging in length from 700 to 910 mm in their natural environment.

In both laboratory and field experiments, some artificial non-eel objects, such as neutrally buoyant wood sticks, were also imaged with the ARIS sonar. The wood sticks are about the same size as the tested eels, which can potentially confuse the identification of eels in a riverine environment. The laboratory experiment is advantageous over the field experiment because it is more controllable, able to reproduce consistent range and velocity values with a known sized object, and needs fewer resources to conduct than field experiments.

The study of transfer learning from lab data to field data serves as an exploration of the feasibility of
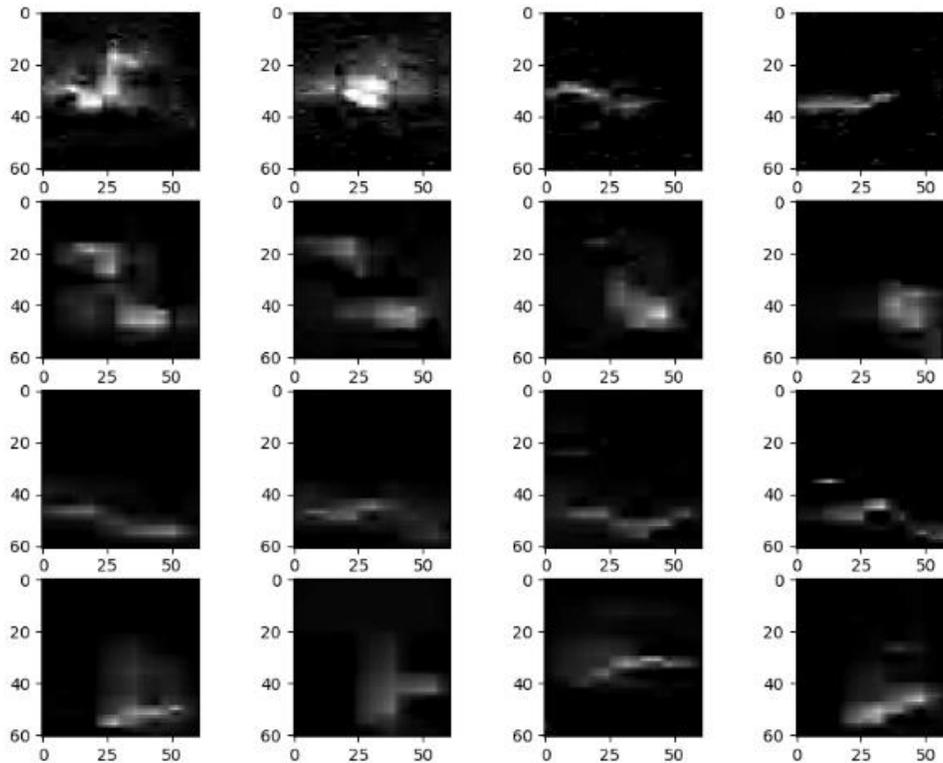
Figure 2. Example images of eels and sticks after background removal and wavelet denoising. First row: eels in the field; second row: sticks in the field; third row: eels in the laboratory; fourth row: sticks in the laboratory.

employing more lab data in the CNN training and minimizing the need and cost of collecting eel data in the field. The eel lengths for the laboratory studies were less than for the adult eels images at the field site. The scaling issue is partially mitigated by the shorter detection range of the sonar. In the laboratory experiment, the detection range of the sonar was set to 2.8-6.7 m, and the eels were imaged at 5.5 m while in the field the range was 2-20 m. Because the apparent size of the object decreases in sonar images when the detection range increases, the lab and field data have similar eel size in the sonar images. Figure 1 includes an example of a raw frame obtained from the ARIS sonar featuring an eel centered at pixel (250, 685).

2.2. Data preprocessing

Multiple data preprocessing techniques are applied to enhance the signal-to-noise ratio of time-lapse sonar images and facilitate object detection. Sonar images sometimes contain not only the object of interest, but also some static structures (such as the pier nose of the dam) that can be treated as image background. Because the background is static and consistent in all frames, it can be removed by image differencing, i.e. by subtracting one image with the mean of several adjacent frames.

In addition to background structures, sonar images can contain unwanted noise (entrained air or small debris) at random location and high intensity. The noise source includes ambient environmental noise that occurs in a similar operating frequency of the sonar Reducing sonar image noise is important for object detection because a large number of high intensity pixels can impede the edge detection of the object. Wavelet analysis can help separate and remove the white noise from the anomalies such as eels, and provide more accurate shapes and dimensions of the objects. Therefore, wavelet denoising was performed after the background has been removed. The denoising process includes (1) computing a wavelet transform of the two-dimensional image and decomposing the image into different frequency components; (2) filtering the wavelet coefficients with a constructed threshold; (3)
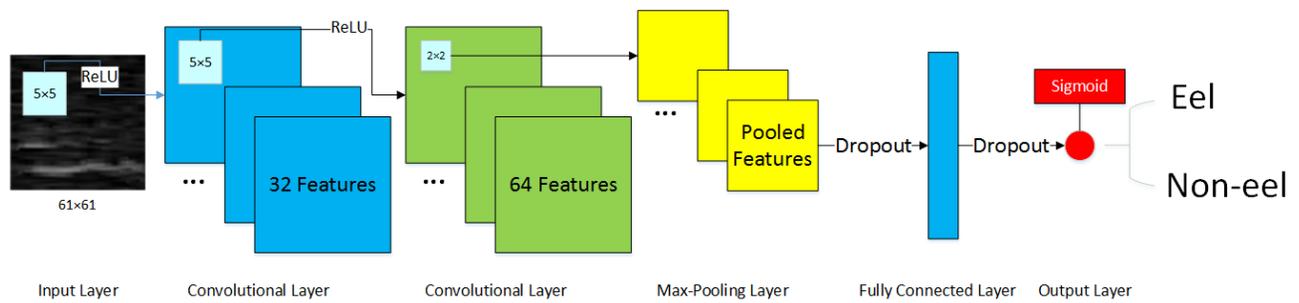
Figure 3. The structure of the convolutional neural network

reconstructing the image using the corresponding inverse wavelet transform [12-14]. Multiple wavelets with various threshold levels were tested and compared to select the right wavelet that can efficiently remove noise from images while maintaining important edges. Based on the results from the comparison, we selected the Daubechies wavelet db2, which is part of the Wavelet Toolbox on MATLAB (Version 2018b, The MathWorks, Natick, MA, USA).

After background removal and wavelet denoising, the image was further processed for object detection. Firstly, a threshold in pixel intensity was selected, and the grayscale image was transformed into a binary image by turning the pixels with intensities above the threshold to 1 and the pixels with intensities below the threshold to 0. Secondly, a sliding window of 61×61 pixels was moved from the top left to bottom right, screening the potential object with a threshold of the number of white pixels (whose pixel intensities equal to 1) in the sliding window. Once the number of white pixels in the sliding window met the threshold and reached the maximum, an object was localized and extracted. Thirdly, the extracted objects were visually verified to ensure that they were either eels or sticks. Note that the extracted image size has been selected to accommodate the eel size in both lab and field data.

Overall, 1,892 eel images and 1,654 stick images were extracted from the lab data, while 129 eel images and 23 stick images were extracted from the field data. In the training of CNN models, we down-sampled the lab data by randomly selecting one out of seven images for both eel and stick images. Some representative images of eels and sticks are included in Figure 1. The eel shapes in the field have more sinusoidal locomotion than those in the lab. Moreover, field stick images appeared as two parts, which was caused by the operation mechanism of the ARIS sonar. In Figure 1.b, the sinusoidal locomotion feature is captured in the four lab eel images as well as the last two field eel images. The first two images of field eels looks different due to its different positioning relative to the sonar beam axis. The four lab stick images are also similar to the last two images of field stick. The first two stick images are cut

into two segments due to sonar imaging mechanism and the motion of sticks. It is obvious that lab data and field data have similarities. The small differences are also notable. The application of transfer learning is built on the big similarities and small differences between the lab and field data. Figure 2 shows example images of eels and sticks after background removal and wavelet denoising.

2.3. Convolutional neural network

In recent years, CNN has achieved remarkable success in various research fields that have a need for image classification. This is due to its many advantages compared to traditional feature-based machine learning [15] including:

1. Automatic feature extraction: the convolutional layers serve as feature extractors that learn features automatically by striding filters (or kernels) through the image data instead of requiring manual feature-engineering.

2. Hierarchical feature extraction: CNN can learn features from the data at different levels, learning both the small details and the big picture.

The convolution function in a CNN is,

$$(f * g)[n] = \sum_{i}^{n} f[i] \times g[i] \qquad (1)$$

where f is the filter feature, g is the input corresponding to the filter, and n is the size of the filter. Such convolution operators introduce a unique property of CNN called parameter sharing. In a traditional neural network, each weight is used for one input unit. Parameter sharing greatly reduces the computational burden compared to dense matrix multiplication.

Parameter sharing also leads to equivariance to input translation, which allows the network to generalize shape patterns like edges and corners in different locations. Moreover, the pooling layer makes the data representation approximately invariant to small translations. These two complementary properties lessen the importance of the exact location of features

[16]. Max pooling especially emphasizes strong features such as edges and corners. Early stage convolutions detect features that are smaller in comparison to convolutions in the deeper layers. Therefore, as the network trains in deeper layers, more complex patterns can be detected.

Activation function is also an important component of neural networks because it makes them non-linear. Such non-linear functions allow modeling complex data distributions. The activation function used in this study was rectified linear unit (ReLU),

$$r(x) = x^+ = \max(0, x) \qquad (2)$$

ReLU handles the vanishing gradient problem well and is computationally less expensive than tanh and sigmoid activation operations [17].

Dropout layers were also incorporated in the CNN structure here. During training, neurons were "turned off" with a static probability, which means that the training process would ignore these neurons with regard to updating weights. Such neurons were "turned on" during the testing period. Dropout layers help the network prevent over-training and generalize better [18]. Batch normalization is usually considered an alternative to dropout. In this study, dropout yielded better results than batch normalization. The output layer used sigmoid function as the classifier,

$$S(x) = \frac{e^x}{e^x + 1} \qquad (3)$$

and binary cross-entropy was used as the loss function,

$$J = -\frac{1}{m} \sum_{j}^{m} \left[y^{(j)} \log(\hat{y}^{(j)}) + (1 - y^{(j)}) \log(\hat{y}^{(j)}) \right] (4)$$

where $\hat{y}$ is the output of the model, y is the true label of the input sample, j stands for the jth sample, and m is the size of the training data.

Applying CNN to distinguish eels from other moving targets (e.g. sticks) is an innovative contribution of this study. Non-eel objects usually have a more rigid shape than eels since the body of eels can twist freely (anguilliform swimming motion). Eels often have a behavior component to their movements. Also, other acoustic characteristics like pixel intensity may be different between the two groups.

The CNN architecture used in this study is shown in Figure 3. After the input layer, there was a convolutional layer with 32 filters. The filter size was 5×5. The second convolutional layer had 64 filters with the same filter size. A max pooling layer followed with 2×2 pooling size. Before and after the fully-connected layer with 128 hidden units, dropouts were implemented. The output layer used the sigmoid function as the final classifier. Adam was chosen as the optimizer [19]. Ten percent of the lab data were separated as the testing set. They were never used by the model during the training
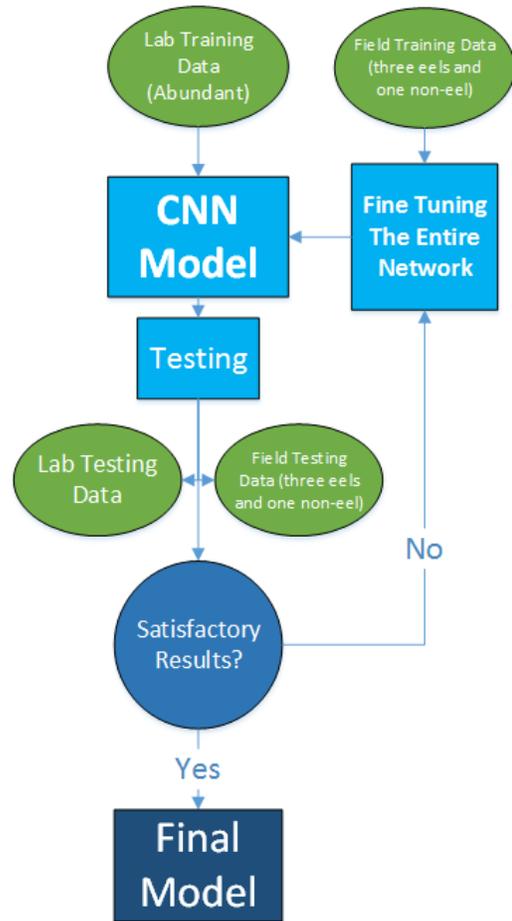

Figure 4. Flowchart of transfer learning

period. For each image, it was normalized by its maximum value.

2.4. Transfer learning

Traditional machine learning assumes that training and testing data must be from the same domain and have the same distribution. However, because deep learning usually requires large amount of training data, transfer learning is becoming more popular in the deep learning community.

Transfer learning is a machine learning framework that transfers knowledge from a certain domain of interest to an application in a related domain. There are typically two reasons for implementing transfer learning:

1. The (large-scale) base model is too hard to train.
2. The target task has limited or insufficient training data.

In both scenarios, the fundamental assumption of transfer learning is that the features learned by the base

model are useful for the target task. The transferability of features in this study comes from the similarity between laboratory and field data. This paper applies transfer learning mainly within the second scenario.

There are several reasons why transfer learning is an attractive option for the target model. There are limited in-river data from field experiment while there are abundant laboratory data available. However, while a model based only on laboratory data might be successful for laboratory environment, the ideal model should be able to identify eels in the field, too. On the other hand, a model trained on the limited field data is very likely to be over-trained and not robust enough.

Transfer learning has been implemented to include field data in the training process and to retain the generalizability of the model based on laboratory data. Moreover, proving the validity of solving the problem of limited field data using transfer learning could potentially lower labor and costs for field experiments which are much more difficult to conduct than laboratory experiments.

The approach (Figure 4) here was to reuse the network that is pre-trained using abundant laboratory data, including its structure and connection weights. We then fine-tuned the weights using field data to obtain a final model that performs well with both testing data sets. The theoretical foundation of this approach is that the extracted features of the base model are versatile and valuable for the final model. The similarities of features include the natural body shapes of eels (edges and corners that can be learned by CNN) in both lab and field, the swimming patterns of eels, the shapes and rigidness of non-eels (such as sticks), and the flowing patterns of non-eels. Both experiments were conducted by domain experts using sonar settings that were mostly the same and in comparable data collection environments in terms of flow velocity, object size, and detection range. The assumption of the proposed approach is that the similarities of features outweigh the differences between the lab and field data.

After fitting a base model that produces good testing results for the laboratory data, the field data were separated into two groups, one for transfer training and one for testing. There were six eel videos from field experiments. The three eel videos that have better image quality were used for fine-tuning the model and the other three lower quality videos were used for testing. By doing so better information was provided to the model while the ability of transfer learning to detect difficult eel cases was also tested. For the two non-eel objects, one was used for transfer training, the other one for testing the fine-tuned model.

The connection weights of the entire network were subject to change during fine-tuning. However, only small updates were expected to be applied to the lab model and a large number of weights could have remained unchanged after transfer learning. The back-propagation process of neural network should be able to find the features that best bridge the gap between the lab and field data and update them accordingly without making huge changes to the entire network. This could be assessed by the classification results of the final model on the lab data. The lab model was fine-tuned using field training data and tested on both lab testing data and field testing data epoch by epoch. The initial learning rate for transfer learning is usually smaller compared to the one used in the lab model training process, since the solution space of the second model is theoretically smaller. This learning rate could also be manually decreased based on the classification performance after each transfer learning epoch.

3. Results and discussion

Two sets of results are summarized in Tables 1 and 2 to illustrate the information gap between laboratory data and field data. These eel classification results are based on all the 129 images from the field data. Table 1 shows the testing results of a model trained on the laboratory data and tested on the field data. The model trained on the lab data can predict the non-eel objects well with correct rates of more than 70%. However, the prediction accuracy on the eels from the field data images is low at 38%.

Table 2 shows the testing results of a model trained on the field data and tested on the laboratory data. This model performed poorly on detecting both eels and non-eel objects. These results indicate an information gap between the lab and field data.

The information gap between lab eel data and field eel data might be mainly due to the different posture and orientation. Eels in lab experiments generally have a more stretched body that is oriented perpendicular to the beam axis of the sonar. The field eels have more posture and orientation variations. When field eels are not oriented perpendicular to the sonar beam axis or are not well stretched as lab eels, their images appear different from lab eel images, as shown in the first two images on the first row of Figure 1. Also, field eel data have a relatively lower resolution because of the longer detection range.

The results of refined transfer learning are summarized in Table 3. The results of field data testing here are object-based for a detailed look at the performance of transfer learning. The images from one video are classified and the whole video is classified based on the percentage of the image classification result. The 129 eel images originally belong to six eel videos and the 23 stick images belong to two non-eel

videos. Without transfer learning, the base model, which was trained on the lab data, achieved a 100% correct classification rate on lab testing data, and also correctly identified the non-eel object if a 50% decision threshold is adopted. However, it performed poorly on field eel data. After two epochs of transfer learning, the number of correctly classified eel images increased, while lab testing results and non-eel field testing results were slightly worse. After five epochs, an ideally tuned model was achieved. All the field eels are correctly identified with a 50% decision threshold. The non-eel object can still be correctly identified. The lab testing result is above 95%, indicating the generalizability of the model.

Results show that the CNN model aided by transfer learning can settle at an ideal middle point between the two data sets. The results support the assumption that the lab and field data are similar with small differences that can be bridged.

The challenges this approach might face could be that the size of the smaller data set is too small compared to the larger one, or that the distributions of the two data sets are too different. In either case, transfer learning might not work as well.

Table 1. Lab model tested on field data

Class	Correct Rate	Percentage
Eels	49/129	38%
Non-eel object 1	9/12	75%
Non-eel object 2	8/11	73%

Table 2. Field model tested on lab data

Class	Correct Rate	Percentage
Eels	94/271	35%
Non-eel objects	69/237	29%

Table 3. Transfer learning results

Class	No TL	2 Epochs	5 Epochs
Lab testing	100%	98%	96%
Field eel 1	10/27	14/27	24/27
Field eel 2	3/17	5/17	10/17
Field eel 3	1/18	4/18	10/18
Non-eel object 2	8/11	7/11	6/11

4. Conclusion

In this study, transfer learning is used to bridge the gap between lab data and field data and a model that works well for both has been developed. This model retains the general representation of eel and non-eel objects from lab data when field data are limited and allows improvement when more field data are included in the model training.

The proposed solution to deal with the challenge of unbalanced data sets could be useful in other fields for utilizing relevant yet different data sets for a balanced model. The automatic eel detection system using sonar data, deep learning, and transfer learning could provide important fish passage monitoring capability for hydropower facility operators concerning safe passage of eels and other species.

6. References

[1] L. Egg, J. Pander, M. Mueller, and J. Geist, "Comparison of sonar-, camera-and net-based methods in detecting riverine fish-movement patterns", Marine and Freshwater Research, 69(12), 2018, pp. 1905-1912.

[2] C. W. Gurshin, D. J. Coughlan, A.-M. Mueller, D. Degan, and P. T. Jacobson, "Eels III: Assessment of Three Sonars to Evaluate the Downstream Migration of American Eels in the St. Lawrence River", The Journal of the Acoustical Society of America, 141, 2017, pp. 3950.

[3] A.-M. Mueller, T. Mulligan and P. K. Withler, "Classifying Sonar Images: Can a Computer-Driven Process Identify Eels?", North American Journal of Fisheries Management, 28:6, 2008, pp.1876-1886.

[4] L. Bothmann, M. Windmann, and G. Kauermann, "Realtime classification of fish in underwater sonar videos", Journal of the Royal Statistical Society: Series C (Applied Statistics), 65(4), 2016, pp. 565-584.

[5] H. Qin, X. Li, J. Liang, Y. Peng, and C. Zhang, "DeepFish: Accurate underwater live fish recognition with a deep architecture", Neurocomputing, 187, 2016, pp. 49-58.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Proceedings of the Neural Information Processing Systems, 2012, pp. 1097-1105.

[7] S.J. Pan, and Q. Yang, "A Survey on Transfer Learning". IEEE Transactions on Knowledge and Data Engineering, 22, 2010, pp. 1345-1359.

[8] D. Zhang, J. Liu, W. Heng, K. Ren, and J. Song, "Transfer Learning with Convolutional Neural Networks for SAR Ship Recognition", IOP Conference Series: Materials Science and Engineering, IOP Publishing, 2018.

[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv preprint arXiv:1409.1556, 2014.

[10] X. Sun, J. Shi, L. Liu, J. Dong, C. Plant, X. Wang and H. Zhou, "Transferring deep knowledge for object recognition in Low-quality underwater videos", Neurocomputing, 275, 2018, pp. 897-908.

[11] EPRI (Electric Power Research Institute), "Assessment of Technologies to Study Downstream Migrating American Eel Approach and Behavior at Iroquois Dam and Beauharnois Power Canal", 3002009406, February 2017.

[12] L. Atallah, and P. Probert Smith, "Using wavelet analysis to classify and segment sonar signals scattered from underwater sea beds", International Journal of Remote Sensing, 24(21), 2003, pp. 4113-4128.

[13] M. R. Azimi-Sadjadi, D. Yao, Q. Huang and G. J. Dobeck, "Underwater target classification using wavelet packets and neural networks", IEEE Transactions on Neural Networks, 11(3), 2000, pp. 784-794.

[14] Z. Hou, Y. V. Makarov, N. A. Samaan and P. V. Etingov, "Standardized Software for Wind Load Forecast Error Analyses and Predictions Based on Wavelet-ARIMA Models -- Applications at Multiple Geographically Distributed Wind Farms", 46th Hawaii International Conference on System Sciences, Wailea, Maui, HI, 2013, pp. 5005-5011.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, 521(7553), 2015, pp. 436-444.

[16] Goodfellow I., Y. Bengio and A. Courville, Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

[17] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines", In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10), Johannes Fürnkranz and Thorsten Joachims (Eds.). Omnipress, USA, 2010, pp. 807-814.

[18] P. Baldi and P.J. Sadowski, "Understanding Dropout", in Proceedings of the Neural Information Processing Systems, 2013.

[19] D. P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization", arXiv:1412.6980v9 [cs.LG] 2017.