



Development of the Deep Learning System for Bird Classification Using Birdsong

Min-Jeong Kang, Yeong-Seon Kim, Hwa-Yeong Shin, Jangwoo Park*

Department of Information & Communication Engineering, Suncheon National University

A B S T R A C T

The activity and distribution of wild birds are biological indicators to evaluate biodiversity. In order to identify bird habitats, collecting and classifying sounds should have to do. Using the bird sound can make easier to distinguish location or type of wild birds. Recently, attempts to analyze bioacoustic data have been risen using the machine learning. We are going to classify the bird songs using deep learning. The bird songs convert into the spectrogram images. Spectrogram images are used for the input of convolutional neural network. In generally the bird song data set for classification contains a lot of noise. Even obtaining the data including noise is difficult. The data is about 200 bird sounds of 20 species. Based on transfer learning, ResNet34, ResNet50 and AlexNet of Convolutional Neural Network are used as the experiment. The experiment parameter is learning rate and epochs. As a result, the ResNet34 shows the highest accuracy of 99.7% and an average of 93% in the test. Therefore, In this paper, we are going to develop the deep learning system that classifies 20 kinds of bird song using ResNet34. By using this system, it can be helpful various activities such as the prevention of avian influenza.

© 2020 KKITS All rights reserved

KEYWORDS : Deep learning, Classification, Spectrogram, AI, Convolutional Neural Networks, ResNet, AlexNet

ARTICLE INFO: Received 27 November 2019, Revised 6 January 2020, Accepted 10 April 2020.

*Corresponding author is with the Department of Information & Communication Engineering, Suncheon National University, 255 Jungang-ro Suncheon, 57922,

KOREA. *E-mail address:* jwpark@suncheon.ac.kr

1. 서론

야생조류 유입 증가로 발생한 많은 질병은 생태계에 매우 큰 영향을 미쳤다. 2014년 1월 야생조류 AI 집단폐사로 인하여 겨울 철새 현황자료 요구가 급증하였다[1]. 야생조류에 대한 질병 및 전염을 예방하기 위해 모니터링을 진행하고 있으며, 야생조류의 서식지와 질병 발생 농가 인근의 접근을 자체시켰다. 하지만 이는 야생조류의 환경 특성상 완벽히 예방할 수 없었다. 따라서 야생조류 서식지의 식별을 위한 효율적이고 확실한 시스템의 개발이 필요한 실정이다. 야생조류 서식지 식별을 위한 시스템 개발에 앞서 새의 종 분류를 위한 시스템 개발이 선행되어야 한다.

최근 머신러닝 및 딥러닝에 관한 관심이 높아지고 있다. 딥러닝은 컴퓨터가 직접 데이터로부터 스스로 학습하는 네트워크 모델로 image, text 등을 직접 분류하고 학습할 수 있다. 이러한 발전은 빅데이터의 활용, 하드웨어의 발전과 인식알고리즘의 개발 등에 힘입어 실용성 높은 기술로 발전되었다. 또한 다양한 형태의 미디어들이 증가하면서 딥러닝의 응용 분야들이 새롭게 창출되고 있다. 그중 1998년에 LeCun 교수에 의해 개발된 CNN(Convolutional Neural Network)은 사람이나 동물의 시각처리 과정을 모방하였다. 역전파 학습알고리즘을 기반으로 영상인식 분야에 성공적으로 적용함으로써 널리 사용되고 있다[2].

본 연구에서는 조류의 소리를 이용해서 조류를 구분할 수 있는 시스템을 개발하고자 한다. 조류의 소리를 바로 딥러닝에 적용할 수 없어서 이미지로 변환하는 과정이 필요하다. 소리를 스펙트로그램(spectrogram) 이미지로 변환하고 CNN에 적용한다. 딥러닝 모델로 사용할 CNN은 뛰어난 성능과 간단한 구조를 가지고 있다. 본 연구에서 분류하고자 하는 새소리 데이터는 환경 특성상 데이터를 구하

기가 매우 힘들고, 한 데이터에서도 여러 노이즈가 포함되어 있어 분류가 더욱 어렵다. 딥러닝 모델로는 AlexNet, ResNet34, ResNet50을 선정하여 성능을 평가하고 조류 분류 시스템에 가장 적합한 모델을 찾는다. 최종 모델로 ResNet34를 선정하였다. 조류의 정보를 제공할 수 있는 웹앱을 구축하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 연구의 이론적 배경에 대해 다룬다. 제3장에서는 개발과정, 제4장에서는 실험결과를 살펴본다. 제5장에서는 결론을 기술한다.

2. 이론적 배경

CNN(Convolutional Neural Network)은 1998년 LeCun이 필기체 인식을 위해 처음으로 고안하였고, Convolution layer, Pooling layer, Fully connected layer 등으로 구성된 심층신경망이다 [3]. 딥러닝에 가장 많이 사용되는 알고리즘 중 하나인 CNN은 모델이 직접 사진, 영상, 글자 또는 소리를 분류할 수 있다. 수많은 은닉계층들이 각 이미지의 서로 다른 특징을 감지한다. 특징을 수동으로 추출할 필요가 없고, 가장 높은 수준의 인식결과를 보여준다[4]. 그리고 복잡하고 고차원적이며 비선형적인 매핑을 학습할 수 있는 능력이 뛰어나다.

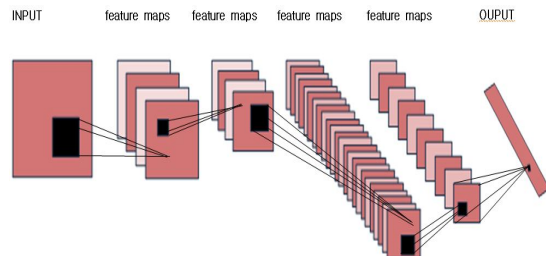


그림 1. CNN의 구조

Figure 1. structure of Convolutional Neural Network

<그림 1>은 CNN의 개략적인 구조를 보여준다. Convolution layer에서 filter가 입력데이터와 Convolution 연산을 한다. Convolution 연산은 filter가 입력데이터에서 stride만큼 돌아다니면서 계산하는 과정이다. filter의 개수만큼 feature map을 얻는다[5]. feature map에 뉴런의 상태를 저장하는 과정이 순차적으로 일어나는데 이는 feature map을 병렬로 실행함으로써 구현된다[3]. feature map은 ReLU에 입력으로 들어간다. ReLU는 CNN에서 많이 쓰이는 Activation function 중 하나이다. ReLU는 복잡한 작업을 처리하기 위해 입력을 비선형적으로 만든다[6]. Activation map이 만들어지고 Pooling layer에 입력된다. Pooling layer는 데이터의 크기를 줄이는 계층이다. 그 방법으로는 max pooling과 average pooling이 있다. max pooling은 최댓값을, average pooling은 평균값을 기준으로 한다[7]. Fully connected layer는 class의 개수만큼 있다. class가 20개면 출력도 20개 나온다. 각각 출력된 class들은 확률값을 가지는데, 가장 큰 값을 가진 class가 최종 예측결과가 된다.

CNN의 대표적인 모델로는 ResNet, AlexNet, VGGNet, GoogLeNet이 있는데, 그 중 ResNet[8]과 AlexNet[9]을 사용하였다. 이 모델들은 각각 2012년, 2015년에 ILSVRC에서 우승하였다. ResNet은 Residual Network의 약자로, CNN의 깊이를 획기적으로 늘린 모델이며 152층의 깊이로 늘어날 수 있었던 원동력은 Residual Block이다. Residual Block에서 중요한 것은 skip connection이다. skip connection은 이전 층에서 넘어온 입력 x 를 중간계층을 거치지 않고 바로 넘겨주어 별도로 연결한 것이다. skip connection의 연산은 아래 식 1로 나타낼 수 있다[10].

$$H(x) = F(x) + x \quad (1)$$

Overfitting이란 학습 과정에서 모델이 훈련 데이터에 과도하게 최적화되어 학습되지 않은 테스트 데이터에 대해서는 잘 예측하지 못하는 현상이다. 훈련 오차가 작음에도 불구하고 실제 데이터에서는 오차가 많이 발생한다[11].

Overfitting이 일어나는 원인은 크게 2가지로 생각할 수 있다. 첫 번째는, 학습에 필요한 데이터가 매우 부족할 경우 발생한다. 이에 대한 해결책은 데이터의 수를 늘리는 것이다. 그러나 상황에 따라 데이터 수집이 더 이상 불가능할 수도 있고, 데이터에 많은 노이즈가 포함되어 있을 수 있다. 두 번째는 훈련 시켜야 할 파라미터가 많아 모델이 복잡해지는 경우에 발생한다. 입력 변수가 많아지면 상대적으로 적은 데이터로 학습되어 데이터의 수가 기하급수적으로 필요해 데이터 부족으로 이어질 수 있다. Overfitting은 Regularization과 Dropout을 사용해서 어느 정도 방지할 수 있다. Regularization은 훈련 데이터와 테스트 데이터의 에러를 줄여 일반화에 적합한 특성을 갖게 만드는 것으로 성능과 모델의 훈련 안전성을 개선하는 효과가 있다. Dropout은 일부 노드만 임의로 사용하여 훈련 데이터에 대한 적합도를 떨어뜨리는 방법이다[12].

3. 개발과정

<그림 2>는 조류 분류 웹앱의 개발과정을 나타내는 순서도이다. 아래에 단계별로 자세한 설명을 하였다.

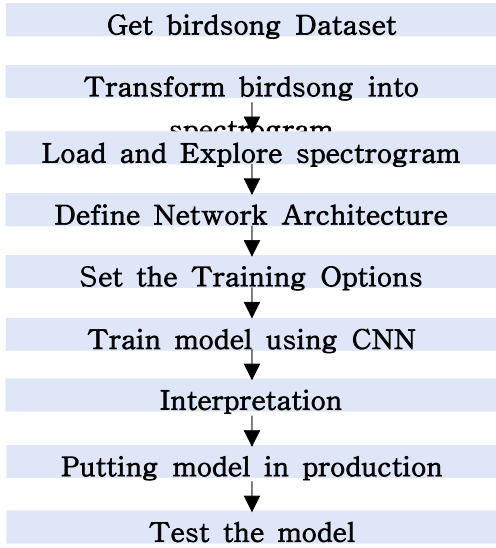


그림 2. 조류 분류 웹앱 개발과정
Figure 2. sequence to develop bird classification web application

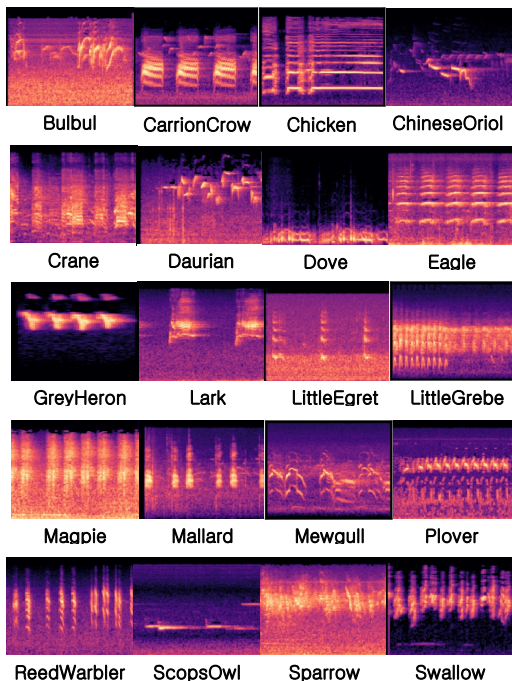


그림 3. 각 클래스의 스펙트로그램
Figure 3. spectrogram for each class

첫째, 실험에 사용된 모든 data는 Youtube, xeno-canto 등에서 수집하였다.

둘째, librosa 라이브러리를 사용하여 소리를 스펙트로그램 이미지로 변환하는 전처리 과정을 거쳤다. 스펙트로그램은 소리의 주파수 스펙트럼을 시각화한 것이다. x축은 시간을 나타내고 y축은 주파수를 나타낸다. 특정 시간의 주파수 진폭을 색상의 농도로 표현한다[13]. <그림 3>은 각 클래스의 스펙트로그램을 나타낸다.

셋째, labeled된 스펙트로그램을 불러서 읽는 단계이다. 입력데이터를 신경망에 고정 크기 입력 계층으로 전송하기 전에 크기를 정규화하고 입력 영역 중앙에 위치시킨다.

넷째, CNN의 모델 중 ResNet34, ResNet50, AlexNet을 사용하였다. class는 20개, 전체 데이터는 2924개이다. validation data는 각 data의 20%로 설정했다. 그리고 test data는 전체 데이터에서 validation data를 제외하고 10개씩 가져왔다. 데이터 크기는 64*64이며 imagenet_stats으로 normalize를 해주었다. <그림 4>는 스펙트로그램을 넣고 은닉계층을 거쳐 출력을 얻는 과정을 개략적으로 보여준다.

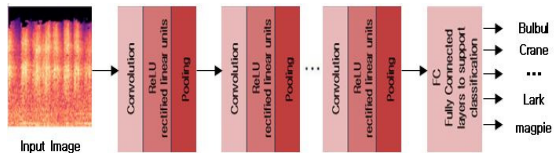


그림 4. 스펙트로그램을 넣은 CNN 동작 과정
Figure 4. CNN process to put the spectrogram

다섯째, overfitting을 방지하기 위해 dropout을 0.5로 설정했다. 그리고 정확도 개선을 기준으로 SaveModelCallback을 사용하여 가장 높은 정확도를 보인 epoch 구간에서 자동으로 저장하였다.

여섯째, 매우 큰 dataset으로 미리 훈련되어있는 모델을 자신의 dataset에 적용해서 사용하는 기술

인 Transfer learning을 사용하였다. 훈련횟수는 50 번과 25번으로 총 6번의 실험을 진행하였으며 learning rate는 $1e-06$, $1e-05$, $1e-04$, $1e-03$, $1e-06 \sim 1e-04$, 음의 기울기가 가장 큰 점으로 실험을 했다. batch size는 16이다.

일곱째, confusion matrix를 그려서 validation data의 predict, actual을 시각화하였다.

여덟째, pkl파일을 생성하고 link generator에서 다운로드 링크를 얻어 모델에 관한 정보를 다운로드 하는 url로 사용하였다. server는 uvicorn, web framework는 starlette를 사용하였다.

마지막으로 local testing server 환경에서 20개의 class를 각 10번씩 총 200번의 테스트를 진행하였다.

4. 실험결과 분석

이전 장에서 제시한 순서도에 따라 조류 분류 연구를 진행하였다. model, 학습횟수, learning rate를 3가지 변수로 설정하여 실험을 수행하고, 조류 분류에 가장 적합한 모델을 찾고자 하였다. <표 1>은 실험 파라미터들이다.

표 1. 파라미터 분석
Table 1. Analysis parameter

Var. #	Name	Range
#1	Model	ResNet34
		ResNet50
		AlexNet
#2	Epochs	25
		50
#3	Learning rate	$1e-06$
		$1e-05$
		$1e-04$
		$1e-03$
		slice($1e-06$, $1e-04$)
		suggested lr

실험에 사용된 데이터는 wav 파일을 스펙트로그램으로 변환하는 전처리 과정을 거쳤다. 20종 각각 200여 개의 소리 파일을 대상으로 하였으며, training data는 2912개, valid data는 728개, test data는 클래스마다 10개씩 총 200개를 사용하였다.

표 2. lr=slice($3e-3$)에서 학습
Table 2. Training at lr=slice($3e-3$)

epochs	accuracy(%)		
	ResNet34	ResNet50	AlexNet
25	93.6	95.0	88.3
50	97.5	97.9	90.2

<표 2>는 각 모델을 기본 learning rate 구간인 slice($3e-3$)에서 25번, 50번 훈련 시켰을 때의 정확도이다[14]. ResNet50에서 가장 높은 정확도를 보였고, AlexNet의 성능이 가장 저조했다. 두 모델은 9.6%의 성능 차를 보인다.

<그림 5>는 validation dataset의 confusion matrix를 나타낸 것이고, 쇠백로(Little Egret)를 다른 새 로 잘못 판단한 경우가 4가지로 가장 많았다.

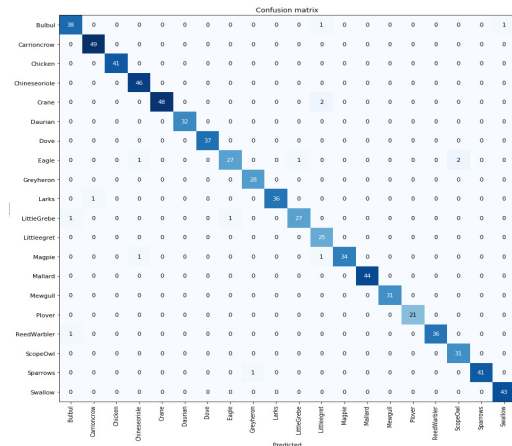


그림 5. ResNet34를 이용한 모델의 confusion matrix
Figure 5. Confusion matrix using of ResNet34

<그림 6>은 학습횟수가 25번인 환경에서 세 가지 모델의 learning rate 변화에 따른 가장 높은 정확도를 나타낸 것이다. 모델(ResNet34, ResNet50, AlexNet)마다 1e-04, 1e-04, 1e-03에서 가장 높은 결과를 얻었고, 그 값은 98.4%, 98.5%, 97%이다.

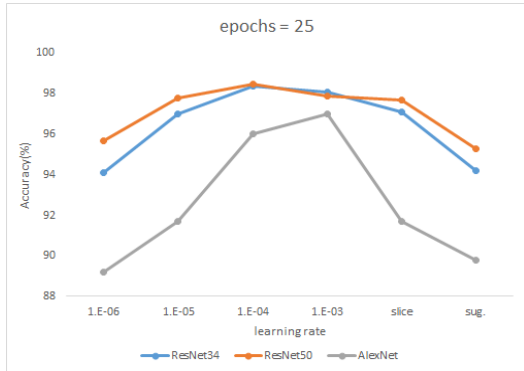


그림 6. 학습횟수가 25번인 모델별 정확도
Figure 6. each model accuracy of epochs=25

<그림 7>은 학습횟수가 50번인 환경에서 세 가지 모델의 learning rate 변화에 따른 가장 높은 정확도를 나타낸 것이다. 모델(ResNet34, ResNet50, AlexNet)마다 1e-03, 1e-04, 1e-04에서 가장 높은 결과를 얻었고, 그 값은 99.7%, 99.6%, 97.6%이다.

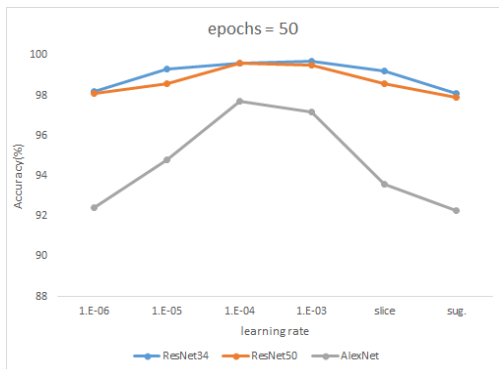


그림 7. 학습횟수가 50번인 모델별 정확도
Figure 7. each model accuracy of epochs=50

정확도가 가장 높은 ResNet34(epochs=50, lr=1e-03)를 production한 후 클래스마다 10번씩 test하였고, <표 3>은 그 결과를 보여준다.

표 3. 테스트 결과
Table 3. Test result

Bird	Accuracy (%)
Bulbul	70
CarrionCrow	100
Chicken	100
ChineseOriole	90
Crane	70
Daurian	80
Dove	100
Eagle	90
Greyheron	90
Lark	90
LittleEgret	90
LittelGrebe	100
Magpie	100
Mallard	100
Mewgull	100
Plover	100
ReedWarbler	100
ScopsOwl	100
Sparrow	100
Swallow	90
Average	93

정확도가 가장 높은 ResNet34(epochs=50, lr=1e-03)를 production한 후 클래스마다 10번씩 test하였고, <표 3>은 그 결과를 보여준다.

학습횟수가 25번인 환경에서, 기본 learning rate=slice(3e-3)으로 모델을 훈련시켰을 때에 비해 ResNet34는 약 4.8%, ResNet50은 약 3.5%, AlexNet은 약 8.7%의 성능 향상을 보였다. 학습횟수가 50번인 환경에서는 각 2.2%, 1.7%, 7.4%의 향상을 보였다. ResNet이 AlexNet보다 validation loss가 낮고 accuracy는 높았다.

6. 결론

본 연구에서는 조류를 분류하기 위해 소리를 스펙트로그램으로 변환하였다. CNN 기반의 ResNet34

를 활용한 조류 분류 시스템을 만들고 조류의 정보를 제공할 수 있는 웹앱을 개발하였다. 조류 분류의 높은 정확도를 얻기 위해 학습횟수와 learning rate 값에 따른 세 개의 모델(AlexNet, ResNet34, ResNet50)로 실험했다. 가장 적합한 모델은 ResNet34임을 확인하였고, test에서 평균 93%의 정확도를 얻었다.

본 연구는 조류의 서식지 식별로 질병 방지 및 지역 환경 변화를 감지하는 수단으로 사용될 수 있고, 생물 다양성 평가와 생태계 모니터링에 활용할 수 있다. 조류 관리체계를 정립하고 더 나아가 멸종 위기종의 보호 및 서식지를 파악하는 기반이 될 수 있다[15]. 조류 전문가 및 생태학자들과 협업을 통하여 더 훌륭한 모델을 개발하는 것이 필요하다.

References

- [1] H-J. Kim, W-H. Hur, J-Y. Park, D-W. Kim, Y-S. Choi, S-G. Kang, J-W. Hwang, and J-Y. Lee, *A winter waterbird monitoring for AI surveillance (2018)*, National Institute of Biological Resources, 2018.
- [2] J-H Park, K-B Hwang, H-M Park, and Y-K Choi, *Application of CNN for fish speies classification*, Journal of the Korea Institute of Information and Communication Engineering, Vol. 23, No. 1, pp. 39-46, 2019.
- [3] Y. LeCun, and Y. Bengio, *Convolutional networks for images, speech, and time series*, The handbook of brain theory and neural networks, Vol. 3361, No. 10, pp. 1-14, 1995.
- [4] What is Convolutional Neural Network? MATLAB & Simulink, <https://kr.mathworks.com/solutions/deep-learning/convolutional-neural-network.html>, Aug. 2019.
- [5] Convolutional Neural Network Abstract, <http://taewan.kim/post/cnn/>, Sep. 2019.
- [6] Activation functions and it's types-Which is better?, <https://towardsdatascience.com/activation-functions-and-its-types-which-is-better-a9a5310cc8f>, Sep. 2019.
- [7] B. J. Son, and K. H. Lee, *Crack recognition of sewer with low resolution using convolutional neural network(CNN) method*, Journal of the Korean Society for Advanced Composite Structures, Vol. 8, No. 4, pp. 58-65, 2017.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
- [9] A. Krizhevsky, I. Sutskever, and G-E. Hinton, *Imagenet classification with deep convolutional neural networks*, In Advances in neural information processing systems, pp. 1097-1105, 2012.
- [10] [ML] Representative CNN models - AlexNet, VGGNet, GoogLeNet, ResNet, <http://blog.naver.com/horajjan/221425497675>, Aug. 2019.
- [11] S-H. Kim, S-J. Oh, G-Y. Yoon, Y-G. Jung, and M-S. Kang, *Influence on overfitting and reliability due to change in training data*, International Journal of Advanced Culture Technology, Vol. 5, No. 2, pp. 82-89, 2017.
- [12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Dropout: a simple way to prevent neural networks from overfitting*, The journal of machine learning research, Vol. 15, No. 1, pp. 1929-1958, 2014.

- [13] E. Browning, R. Gibb, P. G-Kapfer, and K. E. Jones, *Passive acoustic monitoring in ecology and conservation*, WWF-UK, 2017.
- [14] basic_train|fastai,
https://docs.fast.ai/basic_train.html, Oct, 2019.
- [15] W-H. Hur, H-J. Kim, J-W. Hwang, J-Y. Lee, S-G. Kang, D-W. Kim, Y-S. Choi, J-Y. Park, and C-Y. Choi, *A study of bird population monitoring in different habitats (2018)*, National Institute of Biological Resources, 2018.

조류 울음소리를 이용한 조류 분류 딥러닝 시스템 개발

강민정¹, 김영선¹, 신화영¹, 박장우²

¹순천대학교 정보통신공학과 학부생

²순천대학교 정보통신공학과 교수

요 약

야생조류의 활동과 분포는 생물 다양성을 평가할 수 있는 생물학적 지표가 된다. 조류의 서식지를 식별하기 위해서는 소리를 수집하고 분류하는 것이 필요하다. 새소리를 사용하면 야생조류의 위치나 종류를 쉽게 구별할 수 있다. 최근 머신러닝을 이용한 생물음향자료 분석에 대한 시도들이 늘어나고 있다. 딥러닝을 이용해 조류 울음소리를 분류하고자 한다. 조류 울음소리를 스펙트로그램 이미지로 변환한다. 스펙트로그램 이미지는 Convolutional Neural Network의 입력으로 사용한다. 일반적으로 분류하고자 하는 새소리 dataset은 노이즈를 많이 포함하고 있다. 심지어 노이즈를 포함한 데이터조차 구하기 어렵다. 데이터는 20종의 새를 약 200여 개씩 구했다. transfer learning을 기반으로 CNN모델인 ResNet34, ResNet50, AlexNet을 실험에 사용하였다. 실험 파라미터는 learning rate와 학습횟수이다. 그 결과, ResNet34에서 99.7%의 가장 높은 정확도를 보이고, 그 test에서는 평균 93%의 정확도를 보인다. 따라서 본 논문에서는 ResNet34을 이용해 20종의 조류 울음소리를 분류하는 Deep

Learning 시스템을 구현하려고 한다. 이 시스템을 이용하여 조류 인플루엔자의 예방 등 다양한 활동에 도움이 될 수 있다.



Min-Jeong Kang received the bachelor's degree in the Department of Information & Communication Engineering from the Suncheon National University in 2020. Her current research interests include artificial intelligence, statistics and algorithm.

E-mail address: alswj1122kk@naver.com



Young-Seon Kim received the bachelor's degree in the Department of Information & Communication Engineering from the Suncheon National University in 2020. Her current research interests include artificial intelligence and database.

E-mail address: sun505505@naver.com



Hwa-Yeong Shin received the bachelor's degree in the Department of Information & Communication Engineering from the Suncheon National University in 2020. Her current research interests include artificial intelligence and internet of things.

E-mail address: shy970929@naver.com



Jangwoo Park received the B.S., M.S. and Ph.D. degrees in Electronic engineering from Hanyang University, Seoul, Korea in 1987, 1989 and 1993, respectively. In 1995, he joined the faculty member of the Sunchon National University, where he is currently a professor in the Department of Information & Communication engineering. His research focuses on deep learning and machine learning.

E-mail address: jwpark@sunchon.ac.kr