Journal of Korea Robotics Society (2020) 15(2):184-189
https://doi.org/10.7746/jkros.2020.15.2.184 ISSN: 1975-6291 / eISSN: 2287-3961

영상 내 물체 검출 및 분류를 위한 소규모 데이터 확장 기법

Data Augmentation Method of Small Dataset for Object Detection and Classification

김 진 용[1] · 김 은 경[2] · 김 성 신†
Jin Yong Kim[1], Eun Kyeong Kim[2], Sungshin Kim†

Abstract: This paper is a study on data augmentation for small dataset by using deep learning. In case of training a deep learning model for recognition and classification of non-mainstream objects, there is a limit to obtaining a large amount of training data. Therefore, this paper proposes a data augmentation method using perspective transform and image synthesis. In addition, it is necessary to save the object area for all training data to detect the object area. Thus, we devised a way to augment the data and save object regions at the same time. To verify the performance of the augmented data using the proposed method, an experiment was conducted to compare classification accuracy with the augmented data by the traditional method, and transfer learning was used in model learning. As experimental results, the model trained using the proposed method showed higher accuracy than the model trained using the traditional method.

Keywords: Data Augmentation, Deep Learning, Image Synthesis, Perspective Transform, YOLOv2

1. 서 론

최근 그래픽 처리 장치의 발달로 인해 연산 능력이 크게 향상되어, 많은 연산량으로 인해 실시간성의 한계에 부딪혀 암흑기를 맞았던 인공지능 분야가 재조명 받고 있다. 특히 인공지능의 핵심기술 중 하나인 딥러닝에 대한 연구가 활발하게 이루어지고 있으며, 그 중에서도 비전센서를 이용한 영상 데이터 기반의 딥러닝 기술이 화제가 되고있다. 영상 데이터 기반의 딥러닝 기술은 제조 공정의 불량검사, 물체 인식 및 분류 그리고 보안을 위한 홍채 및 지문 인식 등과 같이 다양한 분야에 적용되고있다.

Received : Jan. 29. 2020; Revised : Feb. 20. 2020; Accepted : Mar. 16. 2020
※ This work was supported by BK21PLUS, Creative Human Resource Development Program for IT Convergence and was supported by the Technology Innovation Program(10073147, Development of Robot Manipulation Technology by Using Artificial Intelligence) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea)
1. MS Student, Dept. of Electrical and Computer Engineering, Pusan National University, Busan, Korea (skes1234@pusan.ac.kr)
2. Ph.D. Student, Dept. of Electrical and Computer Engineering, Pusan National University, Busan, Korea (kimeunkyeong@pusan.ac.kr)
† Professor, Corresponding author: School of Electrical and Computer Engineering, Pusan National University, Busan, Korea (sskim@pusan.ac.kr)

딥러닝 기술을 사용하기 위해서는 목적에 맞는 딥러닝 모델을 설계하고, 설계한 모델을 학습하는 단계를 거쳐야한다. 연구되어진 딥러닝 모델 중 CNN (Convolutional Neural Network), R-CNN (Region based Convolutional Neural Network), Fast R-CNN, Faster R-CNN, YOLO (You Only Look Once) 등은 영상 데이터 기반의 물체 인식 및 분류에 특화된 딥러닝 모델이다. CNN 모델은 입력된 영상이 학습된 카테고리 중 어떤 것과 가장 유사한지 계산하여 분류하는 모델로서, 대표적으로 AlexNet, GoogLeNet, VGGNet 등의 유명한 모델들이 사용되어지고 있다[1]. R-CNN, Fast R-CNN, Faster R-CNN, YOLO는 물체의 존재 예상 범위를 추출한 후, 그 범위에 존재하는 물체를 분류한다[2-5]. 영상에 포함된 다중 물체를 분류하고 위치를 파악하기 위해서는 후자의 딥러닝 모델이 주로 사용된다. 딥러닝 모델의 성능을 향상시키려면 학습 데이터가 목적과 관련된 정보를 충분히 나타낼 수 있어야 하는데, 예를 들어 축구공을 학습하기 위해서는 축구공의 다양한 모양, 색깔, 문양 등과 같이 축구공의 특징에 대한 정보를 나타낼 수 있는 학습 데이터가 필요하다. 그리고 딥러닝에 관련된 많은 연구 결과들에 의하면, 학습에 사용되는 데이터의 수가 모델의 성능에 큰 영향을 미친다[6].

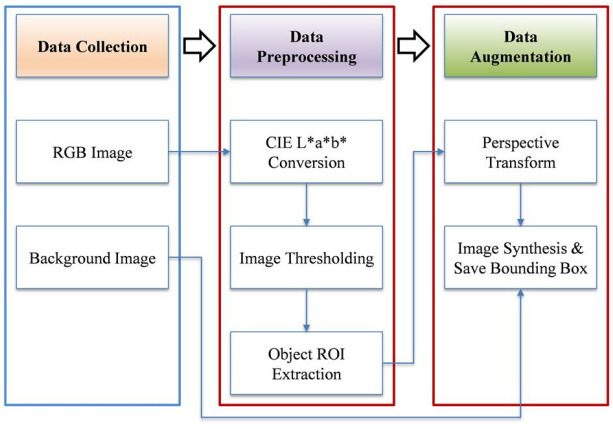
최근 공장의 조립 공정 및 검사 공정 등에서 사용되었던 머신비전 기술이 딥러닝 기술로 대체되고 있는데, 이는 규칙 기반의 머신비전 기술이 해결할 수 없었던 복잡한 문제들을 딥러닝 기술을 이용하여 해결할 수 있기 때문이다[7]. 하지만 딥러닝 기술을 도입하기 위해서는 딥러닝 모델을 학습하기 위한 많은 데이터가 요구된다는 점에서 문제가 발생한다. 공장에서 사용되는 물체는 축구공, 야구공, 풍선 등과 같이 인터넷에서 쉽게 찾을 수 있거나, 데이터 베이스가 구축된 물체가 아니기 때문에 데이터를 얻기 어렵다. 또한 공장에서 사용되는 일부 물체에 대한 데이터는 보안상 공장 관계자 외엔 접할 수도 없고 수집할 수도 없기 때문에, 데이터 구축을 위해 많은 시간과 노력을 투자해야 한다는 한계성이 있다.

따라서 본 논문은 데이터 수집에 한계가 있는 물체를 대상으로 물체 인식 및 분류 그리고 물체 영역 파악을 위한 YOLOv2 모델 학습에 있어서, 최소한의 데이터를 기반으로 투시 변환 및 영상 합성을 이용하여 데이터를 확장하는 기법에 대하여 연구하였다. 본 논문은 2장에서 제안하는 방법에 대하여 설명하며, 3장에서는 제안하는 방법의 성능 검증을 위한 실험 및 결과, 그리고 4장에서는 결론 및 향후 연구에 관하여 얘기하고자 한다.

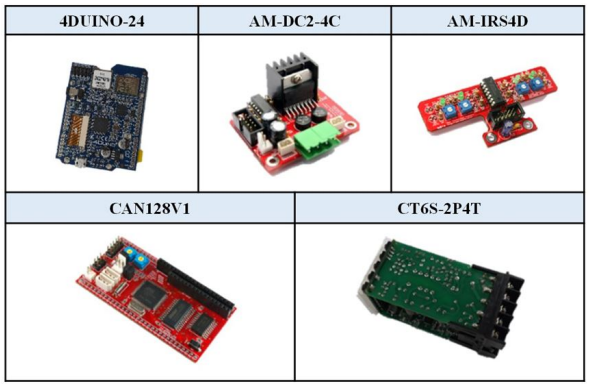
2. 제안하는 방법

본 장에서는 제안하는 방법을 이용하여 소규모 데이터를 기반으로 데이터를 확장시키는 방법에 대하여 설명한다. 제안하는 방법은 크게 데이터 수집, 데이터 전처리, 데이터 변환 및 확장 세단계로 구성되어 있으며, 흐름도는 [Fig. 1]과 같다.

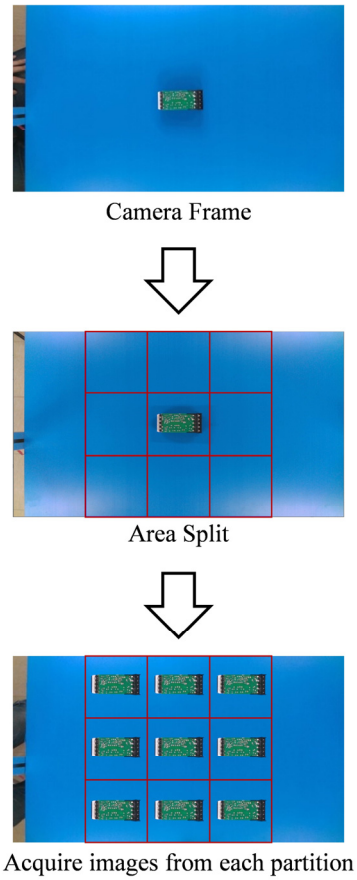
첫 번째 단계인 데이터 수집 단계에서는 데이터 확장을 위한 최소한의 물체 영상을 수집하고, 배경 합성에 쓰일 배경 영상을 수집한다. 물체 인식 대상에 사용될 물체는 데이터 수집에 한계가 있는 5가지 물체를 선정하였으며 [Fig. 2]와 같다. 물체 영상은 비전 센서인 Intel 사의 RealSense D435를 이용하여 위에서 아래로 내려다보는 방향으로 정해진 규칙에 따라 촬영하여 수집하였으며, 그 과정은 [Fig. 3]과 같다. 먼저, 물체와 색상이 겹치지 않는 단색 판넬을 작업대 위에 설치하고, 카메라 프레임의 최대 정방영역을 9등분한 후, 각 등분된 정방영역 위에서 같은 자세로 물체 영상을 수집했다. 또한, 물체가 놓일 수 있는 모든 자세에 대하여 같은 작업을 반복 수행하였으며,



[Fig. 1] Flowchart of the proposed method



[Fig. 2] Selected object for object recognition



[Fig. 3] The process of collecting object images

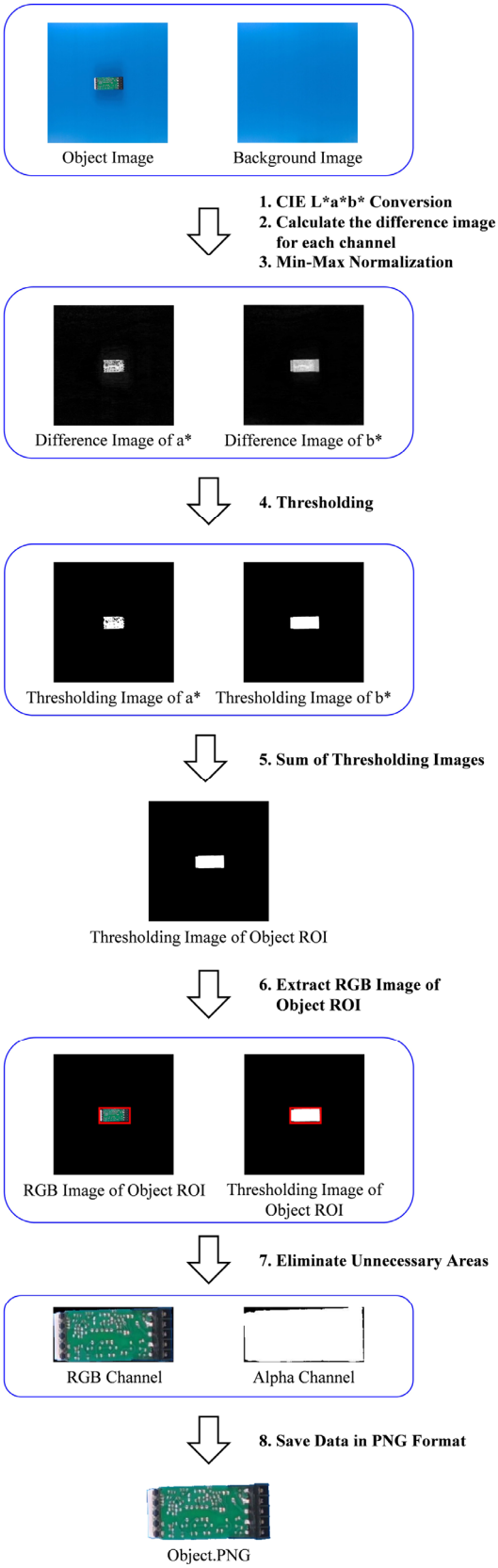
[Table 1] Number of object images

Object Name	Count
4DUINO-24	18
AM-DC2-4C	18
AM-IRS4D	18
CAN128V1	18
CT6S-2P4T	54

이러한 규칙에 따라 수집한 물체 당 데이터 수는 [Table 1]과 같다. 그리고 이후의 전처리인 물체 영역 추출 수행을 위해 단색 판넬 영상을 수집한다. 다음으로, 배경 합성으로 사용될 배경 영상은 물체 인식에 사용할 물체가 포함되어 있지 않은 영상으로 확장할 데이터 수만큼 수집한다. 본 논문은 ILSVRC-2012 Dataset을 사용하였으며, 총 15개의 카테고리에서 각각 1,300장의 영상을 수집하여 총 19,500장의 배경 영상을 수집하였다.

두 번째 단계인 데이터 전처리 단계에서는 [Fig. 4]와 같은 과정을 거쳐, 물체 영상에서 물체에 해당하는 ROI (Region of Interest)의 RGB 영상과 이진 영상을 추출한 후 불필요한 영역을 잘라내고 하나의 파일에 저장하는 단계이다. 먼저, 물체 영상과 단색 판넬 영상에서 프레임의 최대 정방영역을 잘라낸 후, 각각의 영상을 CIE L*a*b* 색공간으로 변환시킨다. 변환된 두 영상의 a*, b*에 해당하는 채널에 대한 차를 구한 후 절대값을 취하게 되면, 차영상은 물체가 존재하는 영역에 대한 픽셀값이 크게 나타난다. 이러한 차영상의 특성을 시각화 하기 위해 a*, b* 채널에 대한 차영상을 0-255 범위의 값으로 정규화한다. 그 다음 정규화 된 차영상의 물체 영역에 대한 픽셀 특성을 분석한 후, 적절한 문턱값을 설정하여 이진화를 수행하고 두 이진 영상을 합하면 물체의 ROI에 해당되는 이진 영상을 얻을 수 있다. 마지막으로 물체 영상에서 물체의 ROI에 해당하는 RGB 영상값을 추출한 후 물체의 ROI를 포함하는 최소 사각 영역을 잘라내는 과정을 거친다. 이렇게 추출한 물체의 RGB 영상과 이진 영상을 한 파일에 저장하기 위해 RGB 채널과 투명도를 결정하는 Alpha 채널을 지원하는 PNG 파일 형식의 특성을 이용하여, 물체의 이진 영상을 Alpha 채널에 저장한다.

세 번째 단계인 데이터 변환 및 확장 단계에서는 투시 변환과 영상 합성을 이용하여 데이터를 확장시키는 단계이다. 일반적으로 데이터를 확장하기 위해 영상을 기하학적 구조를 변환시키는 대칭 변환, 회전 변환, 이동 변환, 전단 변환 등을 포함하는 아핀 변환을 사용한다. 본 논문은 데이터의 다양성을 위해 6-자유도를 가지는 아핀 변환에서 물체의 투영 정도를 결정하는 두 개의 파라미터가 추가한 8-자유도를 가지는 투시 변환을 사용하여 데이터를 확장하였다. 투시 변환식을 동차 좌표계를 이용하여 나타내면 다음과 같다.



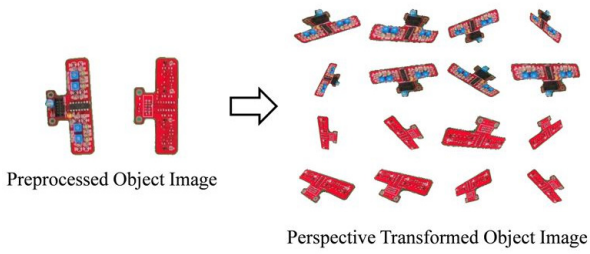
[Fig. 4] The process of extracting the ROI of the object from the object image and saving the ROI of the object

$$\omega\begin{bmatrix}x'\\y'\\1\end{bmatrix}=\begin{bmatrix}h_{11}&h_{12}&h_{13}\\h_{21}&h_{22}&h_{23}\\h_{31}&h_{32}&h_{33}\end{bmatrix}\begin{bmatrix}x\\y\\1\end{bmatrix} \quad (1)$$

식 (1)에서 ω는 동차 좌표계 표현에 있어서의 Scale Factor를 의미하고, x', y'은 변환 후 좌표이며, x, y는 변환 전 좌표이다. 변환 전 좌표 행렬 앞에 곱해져 있는 행렬을 Homography 행렬이라고 부르며, 이동 변환, 회전 변환, 대칭 변환, 투영 변환 등의 변환식을 적절하게 조합하여 구성할 수 있다[8]. 본 논문은 물체의 특성을 최대한 유지하기 위해 대칭 변환, 회전 변환, 투영 변환의 3가지 변환을 이용하여 Homography 행렬을 구성하였다. 대칭 변환에 있어서는 대칭 없음, x축 대칭, y축 대칭 3가지 옵션에 대해 각각 1/3 확률로 선정하여 변환식을 구성하였으며, 회전 변환은 0°-360° 범위에 대하여 Uniform Distribution에 따라 각도를 선정한 후 회전 변환식을 구성하였다. 그리고 투영 변환에 대해서는 물체의 형태를 알아볼 수 있는 정도의 한계값을 실험을 통해 도출하였으며, 0-0.003 범위의 값에 대해 Uniform Distribution에 따라 값을 선정한 후 투영 변환식을 구성하였다. 최종 Homography 행렬은 세 개의 변환식을 곱하여 구성하였으며, 최종 변환식을 이용하여 각 물체당 6,000장의 영상으로 확장하였다. 이러한 투시 변환을 이용하여 데이터를 확장하면 [Fig. 5]와 같은 결과가 도출된다.

마지막으로 [Fig. 6]에서 나타내는 바와 같이 투시 변환을 이용해 확장한 물체 영상과 미리 수집한 배경 영상을 합성하는 동시에 합성한 물체의 영역을 나타내는 Bounding Box 정보를 저장하는 과정을 수행한다. 배경 합성을 하는 이유는 데이터의 복잡성을 증가시킴으로써 딥러닝 모델이 물체와 배경을 분리해내는 능력을 강화하기 위함이다[9]. 먼저, 한 배경 영상에 몇 개의 물체 영상을 합성할지 결정한다. 본 논문에서는 80% 확률로 한 개의 물체, 15% 확률로 두 개의 물체, 5% 확률로 세 개의 물체를 선정하였으며, 물체 선정은 각 카테고리당 20% 확률로 비복원 추출하였다. 그 다음 배경 영상과 물체 영상을 합성하기 위해 물체 영상의 크기를 변환한다. 본 논문에서는 물체 영상을 배경 영상의 15%-40% 비율을 가지도록 Uniform Distribution에 따라 비율을 선정한 후 크기 변환을 수행하였다. 그 다음 배경 영상에 합성할 수 있는 영역에 대해 Discrete Uniform Distribution을 적용하여 선정한 후, 물체 영상의 Alpha 채널에 저장되어 있는 이진 영상에 해당하는 물체 RGB 영상을 배경 영상에 합성한다.

기존의 방법들은 데이터를 확장한 후 확장한 데이터에 대해서 물체가 존재하는 영역에 대한 Bounding Box 정보를 일일이 다 찾아서 저장해야 하는 번거로움이 있어 많은 시간과 노력을 요구한다. 그러나, 제안하는 방법은 물체의 ROI에 대한



[Fig. 5] The result of performing perspective transform on the preprocessed images

[Fig. 6] Result of displaying bounding box after combining object image and background image

[Table 2] The number of times used for extended image data for each object

Object Name	Count
4DUINO-24	4,742
AM-DC2-4C	4,969
AM-IRS4D	4,780
CAN128V1	4,857
CT6S-2P4T	4,928

정보와 합성 위치에 대한 정보를 알고 있기 때문에, 영상 데이터 확장과 동시에 영상 내의 물체의 Bounding Box 정보를 같이 저장할 수 있어, 많은 시간을 절약할 수 있다는 장점이 있다. 이에 본 논문은 이러한 이점을 이용하여 미리 수집한 배경 영상을 비복원 추출하여 합성하는 방법을 활용하여 최종적으로 126장의 학습데이터를 19,500장의 학습 데이터로 확장하였다.

물체 영상 데이터 수의 평형성을 위해 각 물체 영상의 사용된 수를 체크한 결과는 [Table 2]에서 나타냈으며, 5개의 물체 영상의 사용 빈도가 비슷함을 확인할 수 있었다.

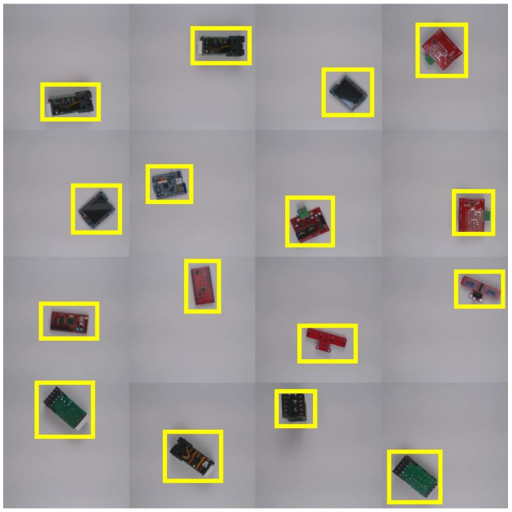
3. 실험 및 결과

제안하는 방법의 검증을 위해 기존의 모델을 전이학습한 후 학습에 사용되지 않은 테스트 데이터에 대한 분류 정확도를 비교하는 실험을 진행하였다.

제안하는 방법을 이용하여 확장한 데이터가 딥러닝 모델 성능 향상에 효과적임을 검증하기 위해 [Table 1]에 나와 있는 원본 물체 영상 데이터를 기반으로 기존의 방법인 아핀 변환을 사용하여 확장한 학습 데이터를 비교용 학습 데이터로 사용하였으며, 물체 영역에 대한 정보는 수동적으로 Bounding Box을 마크하여 YOLOv2 모델 학습을 위한 데이터 타입으로 제작하였다. 학습 모델은 분류 모델을 Inception-v3 모델로 구성한 YOLOv2 모델을 이용하였고, 마지막 분류층의 Fully Connected Layer를 5개의 노드를 가지는 Fully Connected Layer로 대체하였다[10]. 이 모델을 비교용 학습 데이터와 제안하는 방법으로 확장한 데이터를 각각 전이학습 하였으며, 학습 옵션은 [Table 3]에 나타낸 바와 같다. 비교 실험의 형평성을 위해서는 두 학습 데이터군의 데이터 수를 맞춰야 하지만, 앞에서 언급했듯이 기존의 방법을 이용한 데이터 확장기법은 확장한 모든 학습 데이터에 대해 Bounding Box를 마크해야 되므로 19,500장의 학습 데이터로 확장하는 데에 한계가 있다. 이에 본 논문은 원본 데이터를 학습할 시 MATLAB의 데이터 확장 함수를 이용하여 매 반복마다 학습 데이터에 대하여 아핀 변환을 수행하는 방법을 이용하여 학습하였다.

[Table 3] Training options for learning deep learning models

Option	Value
Max Epoch	100
Mini Batch Size	32
Learning Rate	0.001
Learning Rate Reduction Cycle	Every 20 Epoch, Learning Rate×0.8
L2 Regularization	0.0001
Optimization Function	SGDM Momentum = 0.9



[Fig. 7] Test data for verifying the proposed method

[Table 4] Results of classifying accuracy of test data and training data for each model

	Test Data	Training Data
Traditional Method	9.2%	85.7%
Proposed Method	**86.8%**	99.8%

학습된 두 모델의 성능 비교를 위해 [Fig. 7]과 같이 학습에 사용되지 않은 테스트 데이터를 구성하여 물체 인식 정확도를 도출하는 실험을 수행하였고, 두 모델의 과적합 여부 판별을 위해 학습 데이터에 대한 물체 인식 정확도 도출 실험 또한 수행하였다. 실험 결과는 [Table 4]와 같다. 먼저, 학습에 사용되지 않은 테스트 데이터에 대해서는 제안하는 방법으로 데이터를 확장하여 학습한 모델의 성능이 86.8%로, 기존의 방법으로 데이터를 확장하여 학습한 모델의 성능인 9.2%보다 더 높은 분류 정확도를 보였다. 다음으로, 과적합 여부 판단을 위해 학습 데이터로 실험한 결과는 기존의 방법으로 확장한 데이터로 학습한 모델은 테스트 데이터에 대해서는 9.2%의 낮은 정확도를 보였음에 반해, 학습 데이터에 대해서는 85.7%의 높은 정확도를 보였다. 이는 기존의 방법을 이용하여 확장한 데이터로 학습한 모델이 학습 데이터에 과적합 되었다고 볼 수 있다.

그에 반해 제안하는 방법은 테스트 데이터에 대한 성능과 학습 데이터에 대한 성능이 크게 차이가 없다는 것을 확인할 수 있다. 따라서 제안하는 방법을 이용하여 최소한의 소규모 데이터를 기반으로 확장한 데이터가 기존의 방법에 비해서 YOLOv2 모델 학습에 유효하다는 것과, 과적합을 피하기에 충분한 데이터 다양성을 가지고 있다는 것을 검증하였다.

4. 결론 및 향후 연구

실험 결과, 제안하는 방법을 이용하여 데이터를 확장시킨 후 YOLOv2 모델을 학습시켰을 때, 기존의 방법을 이용하여 데이터를 확장시킨 후 학습시킨 모델보다 테스트 영상의 분류 정확도가 높았으며, 과적합 문제 및 데이터 다양성 문제 또한 해결할 수 있었다.

향후 연구로 제안하는 방법의 전처리 과정 중, Alpha 채널이 물체에 해당하는 픽셀 정보를 가지고 있다는 점을 이용하여 Mask-CNN 모델을 대상으로 하는 학습 데이터를 확장시키는 기법에 대하여 연구하고자 한다.

References

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541-551, 1989, DOI: 10.1162/neco.1989.1.4.541.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587, Columbus, OH, USA, 2014, DOI: 10.1109/CVPR.2014.81.

[3] R. Girshick, "Fast r-cnn," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440-1448, 2015, DOI: 10.1109/ICCV.2015.169.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, June, 2015, DOI: 10.1109/TPAMI.2016.2577031.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779-788, 2016, DOI: 10.1109/CVPR.2016.91.

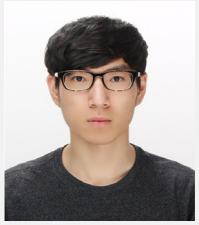
[6] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv: 1712.04621*, 2017, [Online], https://arxiv.org/pdf/1712.04621.pdf.

[7] D. H. Kim, *Manufacturing Transition: Smart Factory*, [Online], https://brunch.co.kr/@duk-hyun/44, Accessed: Jan. 26, 2019.

[8] E. Dubrofsky, "Homography estimation," B. S. thesis dissertation, THE UNIVERSITY OF BRITISH COLUMBIA, Vancouver, Canada, 2009, [Online], https://www.cs.ubc.ca/grads/resources/thesis/May09/Dubrofsky_Elan.pdf.

[9] J. Y. Kim, "Performance Improvement of Deep Learning using Data Augmentation based on Perspective Transform and Image Synthesis," M.S thesis, Pusan National University, Busan, Korea, 2020, [Online], http://www.dcollection.net/handler/pusan/000000143862.

[10] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285-1298, May, 2016, DOI: 10.1109/TMI.2016.2528162.



김 진 용

2018 부산대학교 전기공학과(공학사)

2018~현재 부산대학교 전기전자컴퓨터공학과 석사과정

관심분야: Mobile Robot, Deep Learning

김 은 경

2014 부산대학교 전자전기공학부(공학사)

2016 부산대학교 전자전기컴퓨터공학과 (공학석사)

2016~현재 부산대학교 전기전자컴퓨터공학과 박사과정

관심분야: Intelligent System, Object Recognition, Robot Vision, Image Processing

김 성 신

1986 연세대학교 전기공학과(공학석사)

1996 Georgia Inst. of Technology, 전기 및 컴퓨터공학부(공학박사)

1998~현재 부산대학교 전기컴퓨터공학부 교수

관심분야: Intelligent System, Intelligent Robot, Fault Diagnosis and Prediction