The Journal of the Convergence on Culture Technology (JCCT)
Vol. 10, No. 3, pp.909-914, May 30, 2024. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2024.10.3.909

JCCT 2024-5-104

로봇 비전의 영상 인식 AI를 위한 전이학습 정량 평가

Quantitative evaluation of transfer learning for image recognition AI of robot vision

정재학*

Jae-Hak Jeong*

요 약 본 연구에서는 로봇 비전용 영상 인식을 비롯한 다양한 AI 분야에서 널리 활용되는 전이학습에 대한 정량적 평가를 제시하였다. 전이학습을 적용한 연구 결과에 대한 정량적, 정성적 분석은 제시되나, 전이학습 자체에 대해서는 논의되지 않는다. 따라서 본 연구에서는 전이학습 자체에 대한 정량적 평가를 숫자 손글씨 데이터베이스인 MNIST를 기반으로 제안한다. 기준 네트워크를 대상으로 전이학습 동결층의 깊이 및 전이학습 데이터와 사전 학습 데이터의 비율에 따른 정확도 변화를 추적하였다. 이를 통해 첫번째 레이어까지 동결할 때 전이학습 데이터의 비율이 3% 이상일 경우, 90% 이상의 정확도를 안정적으로 유지할 수 있음이 확인되었다. 본 연구의 전이학습 정량 평가 방법은 향후 네트워크 구조와 데이터의 종류에 따라 최적화된 전이학습을 구현하는데 활용 가능하며, 다양한 환경에서 로봇 비전 및 이미지 분석 AI의 활용 범위를 확대할 것이다.

주요어 : 로봇 비전, 영상 인식, 전이 학습, 합성곱신경망, MNIST

Abstract This study suggests a quantitative evaluation of transfer learning, which is widely used in various AI fields, including image recognition for robot vision. Quantitative and qualitative analyses of results applying transfer learning are presented, but transfer learning itself is not discussed. Therefore, this study proposes a quantitative evaluation of transfer learning itself based on MNIST, a handwritten digit database. For the reference network, the change in recognition accuracy according to the depth of the transfer learning frozen layer and the ratio of transfer learning data and pre-training data is tracked. It is observed that when freezing up to the first layer and the ratio of transfer learning data is more than 3%, the recognition accuracy of more than 90% can be stably maintained. The transfer learning quantitative evaluation method of this study can be used to implement transfer learning optimized according to the network structure and type of data in the future, and will expand the scope of the use of robot vision and image analysis AI in various environments.

Key words : Robot Vision, Image Recognition, Transfer Learning, Convolution Neural Network, MNIST

Ⅰ. 서 론

로보틱스 분야에서 가장 널리 활용되는 센서는 카메라를 이용한 비전 (vision)이며, 환경 및 물체 인식, 자기 위치 파악, 거리 측정 등 다양한 용도를 위한 영상 인식 (image recognition)이 활발히 연구되고 있다[1].

*정회원, KAIST 기계공학과 박사 (제1저자, 교신저자)
접수일: 2024년 3월 19일, 수정완료일: 2024년 4월 25일
게재확정일: 2024년 5월 10일

Received: March 19, 2024 / Revised: April 25, 2024
Accepted: May 10, 2024
*Corresponding Author: jaehak.jeong@kaist.ac.kr
Dept. of Mechanical Engineering, KAIST, Korea

영상 인식 분야에서 딥러닝 (deep learning)을 비롯한 인공지능 (AI, artificial intelligence)의 도입으로 그 응용 범위와 성능이 폭발적으로 확장되고 있다[2]. 이러한 영상 인식을 비롯한 AI 분야에서, 가장 학습에 중요한 것은 대량의, 균질한 데이터이다[3]. ImageNet, CIFAR-10 등의 공개 이미지 데이터베이스가 널리 활용되고 있으며, 이러한 정제된 데이터베이스는 양이 방대하고 레이블 (label) 분포가 균일하여 빠르고 정확한 학습의 기반이 된다. 그러나, 영상 인식 AI의 응용 목표인 특정 물체 인식, 의료 영상 진단, 이상 상태 감지 등의 분야는 이러한 공개 데이터를 활용할 수 없으며, 직접적으로 취득한 소량의, 불균일한 데이터를 기반으로 한다[4]. 이를 극복하기 위하여 데이터 증강 (data augmentation), 비지도 학습 (non-supervised learning) 등의 방법이 제시되었다[5].

전이학습 (transfer learning)은 학습 데이터의 양과 질 부족을 극복하기 위한 방법으로, 목표 domain(영역)과 유사한 domain의 데이터를 확보하여 사전학습 후, 소량의 목표 domain 데이터를 이용하여 재학습 하는 영역적응(domain adaptation) 방법의 일종이다[6, 7]. 전이학습은 영상 인식을 비롯한 다양한 AI관련 연구에서 활발히 응용되나, 전이학습 그 자체에 대한 분석 없이 그 결과에 대한 정성적, 정량적 분석과 평가만 이루어진다[6, 7].

본 연구는 전이학습 자체에 대한 정량 평가 방법을 제시하기 위하여, MNIST 및 TMNIST 데이터베이스를 이용한다. 이를 기반으로 전이학습에서 네트워크 동결 레이어 깊이, 사전훈련 대비 전이학습 데이터 비율 등 전이학습 조건이 인식 정확도에 미치는 영향에 대한 정량 평가를 수행하였다.

II. 학습 데이터베이스

2.1. MNIST 데이터베이스

본 연구에서는 영상 인식 AI의 성능 평가에 널리 활용되는 MNIST (Modified National Institute of Standards and Technology) 데이터베이스를 활용한다[8]. MNIST는 손글씨로 적은 70000개의 0-9까지의 숫자 이미지 데이터로, 28x28 픽셀로 균등한 크기와 안티앨리어싱, 그레이스케일 레벨 0-225 전처리가 완료되어 있다.

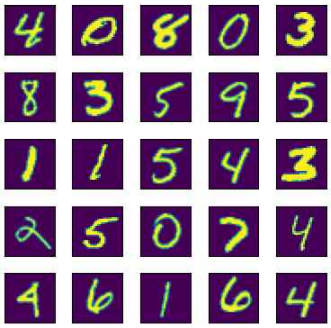

그림 1. MNIST 데이터베이스 샘플 이미지
Figure 1. Sample images of MNIST database.

그림 1은 MNIST의 샘플 이미지를 일부 나타낸 것이다. 손글씨 특유의 왜곡과 개인차로 인해 사람도 혼동될만큼 어려운 데이터가 포함된다. 이처럼 MNIST는 실제적인 오류와 왜곡을 포함하며, 때문에 본 연구의 목표인 전이학습 및 그 정량 평가에 용이하다.

본 연구에서는 방대한 MNIST 데이터베이스의 일부만을 활용하여 전이학습 데이터로 활용한다. 이는 데이터 취득의 현실적인 한계를 반영하기 위하여 그 양을 제한하였으며, 후에 비교할 TMNIST 데이터베이스 대비 1%인 300개의 소량 데이터만 활용된다.

2.2. TMNIST 데이터베이스

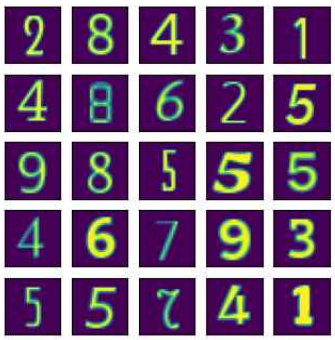

그림 2. TMNIST 데이터베이스 샘플 이미지
Figure 2. Sample images of TMNIST database.

또한, 본 연구에서는 영상 인식 AI의 사전 훈련 (pre-training)에 TMNIST (Typeface MNIST) 데이터베이스를 활용한다[9]. TMNIST는 컴퓨터로 생성된 2990개의 디지털 폰트(서체) 기반의 29900개의 0-9까지의 숫자 이미지 데이터로, MNIST와 동일하게 28x28 픽셀로 균등한 크기와 안티앨리어싱, 그레이스케일 레

벨 0-225 전처리가 완료되어있다. 그림 2는 TMNIST의 샘플 이미지를 일부 나타낸 것이다. 손글씨와 유사하게 다양한 디지털 폰트의 형태로 변조되어 있으면서도, 숫자들의 각각의 특징을 뚜렷이 확인할 수 있다.

본 연구에서는 기준 네트워크의 사전 훈련에 TMNIST 데이터베이스 전체를 활용한다. 통제된 환경에서, 균등한 분포를 갖도록 생성된 TMNIST 데이터베이스는 각 숫자의 특징점을 빠르게 학습하는데 유리하다. MNIST의 개인차와 동일하게 2990개의 디지털 폰트의 다양성을 내포하므로, MNIST에 대한 전이학습 적용이 가능하다.

III. 영상 인식 AI를 위한 전이학습

3.1 기준 네트워크 모델

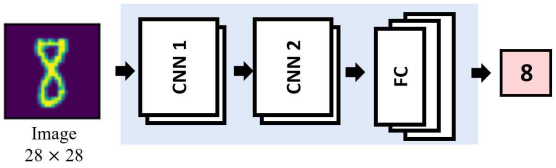

그림 3. 기준 네트워크의 2중 CNN 구조
Figure 4. 2-layer CNN structure of reference network.

본 연구에서는 영상 인식을 위한 AI 모델의 기준 네트워크를 2중 CNN (합성곱신경망, Convolution Neural Network) 구조로 구성하였다. 입력은 28x28 픽셀의 이미지로 데이터이며, 2중의 convolution layer와 1층의 fully connected layer를 거쳐 숫자를 인식한다. 각각의 convolution layer 는 2D convolution (Kernel size 3), ReLU, Maxpooling (Kernal size 2), Dropout (50%) 층을 포함하여 빠른 수렴과 과적합을 예방한다. 마지막에 Fully Connected Layer를 이용해 0-9 사이의 숫자로 분류, 인식한다.

그림 3은 기준 네트워크의 구조를 나타낸다. 네트워크는 PyTorch를 이용하여 구현되었으며. Batch Size는 100을 사용하였다. 학습은 10 Epoch만 수행되었으며, 학습 속도와 정확도를 평가하였다. 훈련과 평가를 위해 입력데이터는 80%의 train set, 20%의 test set으로 나누어 사용되었다.

3.2 사전 학습 결과

전이학습의 필요성을 평가하기 위하여, 소량의 300개 MNIST 데이터 만으로 기준 네트워크의 학습을 수행하였다. 0-9 비율을 동일하게 240개의 train set과 60개의 test set으로 구분하였으며, 240개의 train set로 기준 네트워크를 훈련하였다.

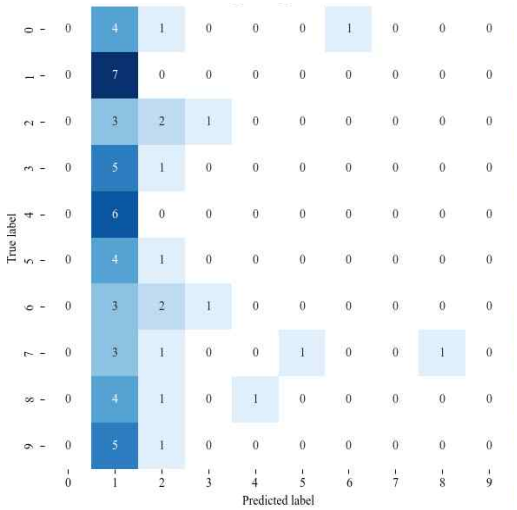

그림 4. 300개 MNIST 데이터로 훈련된 기준 네트워크 인식 결과
Figure 4. Recognition result of reference network trained with 300 of MNIST data

그림 4는 훈련된 기준 네트워크를 이용하여 60개의 test set을 인식한 결과를 혼동행렬로 나타내었다. 300개의 소량의 MNIST 데이터만 이용할 경우, 학습 데이터의 부족으로 정상적으로 훈련이 되지 않으며 15%의 인식 정확도를 보였다. 이처럼 현실적으로 확보 가능한 소량의 데이터 뿐만으로는 영상 인식 AI 모델의 정상적인 훈련이 어려움이 확인되었다.

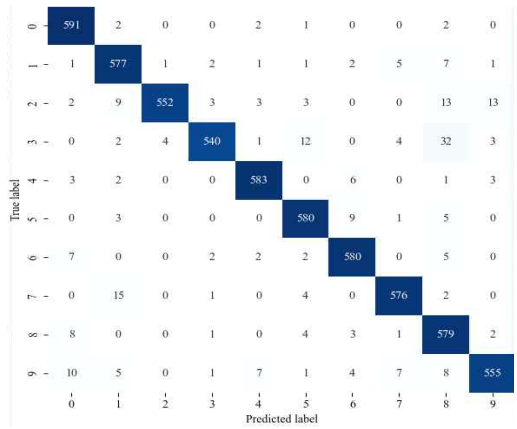

그림 5. 29900개 TMNIST 데이터로 훈련된 기준 네트워크 인식 결과
Figure 5. Recognition result of reference network trained with 29900 of TMNIST data

전이학습을 위한 사전훈련을 기준 네트워크에 적용하기 위하여, 대량의 29900개의 TMNIST 데이터를 이용하였다. 0-9 비율을 동일하게 23920개의 train set과 5980개의 test set으로 구분하였으며, 23920개의 train set로 기준 네트워크의 사전훈련을 수행하였다.

그림 5는 사전훈련된 기준 네트워크를 이용하여 5980개의 test set를 인식한 결과를 분류한 결과를 혼동행렬로 나타내었다. 그 결과, 10 epoch만에 96.77%의 인식 정확도를 확보하였다. 즉 TMNIST는 대량의, 균질하고 우수한 품질의 데이터베이스이며, 이를 통해 빠르게, 높은 정확도로 사전학습이 가능하였다.

3.3. 전이학습 구현

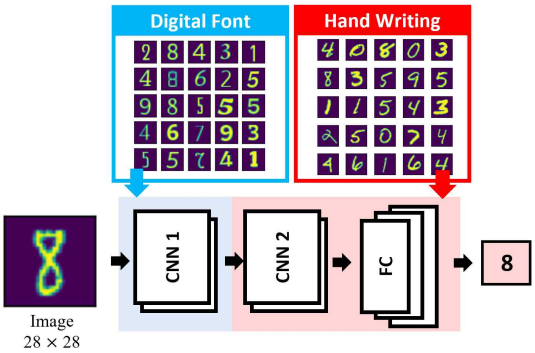

그림 6. 전이학습 적용 개념도
Figure 6. Schematic of applying transfer learning.

전이학습은 사전훈련된 네트워크에 소량의 전이학습 데이터를 이용하여 부분적으로 추가 학습을 진행하는 방법으로, 도메인 전이 (domain adaptation) 기법의 일종이다. 본 연구의 MNIST 및 TMNIST와 같이, 유사하되 다른 특징, 분포를 갖는 경우에 적합하다.

그림 6은 이를 본 연구의 전이학습 적용 개념도이다. 본 연구에서는 전이학습을 대량의 TMNIST로 사전훈련된 기준 네트워크에서, 1번째 layer의 gradient는 동결(freeze)하고, 2번째 layer 이후의 네트워크의 gradient를 소량의 MNIST 데이터를 이용하여 업데이트하는 방식으로 구현하였다. 전반부는 TMNIST 데이터베이스로 학습되어 숫자의 특징을 추출하는 부분으로, 이후 전이학습 도중 바뀌지 않도록 동결하였다. 후반부는 1번째 layer에서 추출된 특징을 기반으로 0-9 사이의 숫자를 인식을 담당하며, 전이학습을 진행하며 MNIST 데이터로 학습되며 업데이트되어, 보다 MNIST에 적합하도록 전이학습이 적용되었다.

3.4 전이학습 결과

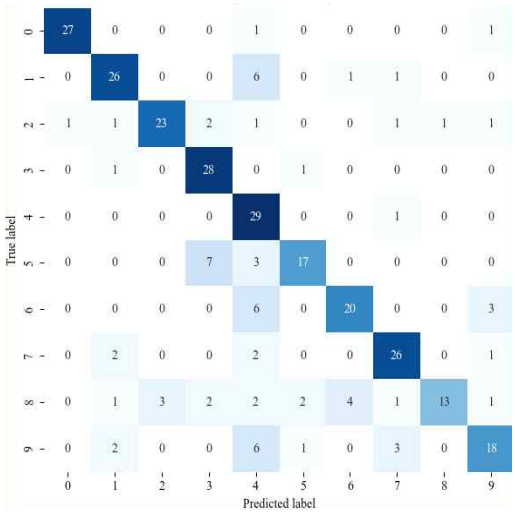

그림 7. 전이학습 적용 이전, TMNIST 데이터로 사전 훈련된 기준 네트워크의 MNIST 데이터 인식 결과
Figure 7. Before applying transfer learning, the result of recognizing MNIST data using a reference network pre-trained with TMNIST data.

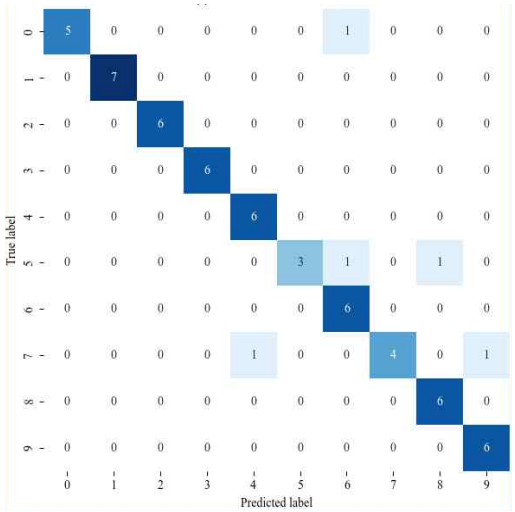

그림 8. 전이학습 적용 이후, TMNIST 데이터로 사전 훈련되고 MNIST 데이터로 업데이트된 기준 네트워크의 MNIST 데이터 인식 결과
Figure 8. After applying transfer learning, the result of recognizing MNIST data using a reference network pre-trained with TMNIST data and updated with MNIST data.

전이학습을 적용하기 전에, TMNIST로 사전훈련만 적용된 상태에서 기준 네트워크의 MNIST 인식 정확도를 평가하였다. 그림 7은 29900개의 TMNIST 데이터로 훈련된 기준 네트워크가 300개의 MNIST 데이터를 인식한 결과를 혼동행렬로 나타낸 것이다. 인식 정확도는

75.67%로, 혼동행렬을 통해 숫자 5, 8, 9에 대한 오류가 다수 발생함이 확인되었다. 이는 명확하게 숫자를 구분할 수 있는 디지털 서체 기반 TMNIST와 달리, 손글씨인 MNIST는 개인차에 의한 왜곡이 존재하며, 다른 숫자로 혼동될 여지가 있다.

전이학습을 적용한 이후, 분류 정확도를 평가하였다. 전이학습에는 MNIST 데이터를 이전과 동일하게 240개의 train set과 60개의 test set으로 구분하여 사용하였으며, 240개의 train set로 사전훈련된 기준 네트워크의 후반부 레이어 만을 학습시켰다.

그림 8은 전이학습이 적용된 이후 기준 네트워크가 60개의 MNIST test set에 대하여 인식을 수행한 결과를 혼동행렬로 나타낸 것이다. 인식 정확도는 91.67%로 전이학습 적용 이전에 비하여 대폭 상향되었으며, 혼동행렬에서 특정 숫자에 집중되어 있었던 오류들이 사라졌다.

Ⅳ. 전이학습 정량 평가

앞서 대량의 TMNIST 데이터로 사전훈련된 기준 네트워크에 소량의 MNIST 데이터를 이용해 전이학습을 수행하는 것으로 그 인식 정확도를 대폭 향상 시킬 수 있음을 확인하였다. 본 연구에서는 전이학습 조건 자체에 대한 정량평가를 수행하기 위하여 전이학습 파라미터에 따른 인식 정확도를 추적한다. 조절 가능한 전이학습 파라미터는 동결하는 레이어의 깊이, 사전 훈련 데이터와 전이학습 훈련 데이터의 비율을 고려할 수 있다. 본 연구에서 활용한 전이학습 파라미터 및 조절 범위를 표 1에 나타내었다.

표 1. 전이학습 파라미터 조절 범위
Table 1. Transfer learning parameters range

Layer Freeze Depth	Freeze after 1st Convolution Layer
	Freeze after 2nd Convolution Layer
Ratio of Data Transfered	0.5 % - 10 %

3.2 사전 학습 결과

본 연구에서는 전이학습 자체의 대한 정량적 평가를 수행하기 위하여 표 1과 같이 전이학습 파라미터를 조절하며 인식 정확도를 추적하였다. 동결 레이어 깊이를 첫번째, 또는 두번째 layer까지 지정한 경우의 인식 정확도 변화 경향을 모니터링한다. 또한 사전 훈련에 사용되는 TMNIST 데이터 대비 전이학습에 사용되는 MNIST 데이터의 비율을 조절하며 인식 정확도의 변화를 측정하였다.

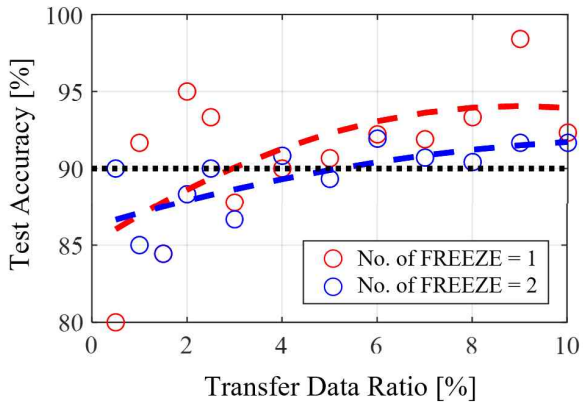

그림 9. 전이학습 조건에 따른 인식 정확도 변화
Figure 9. Changes in recognition accuracy according to transfer learning conditions.

그림 9는 전이학습 파라미터에 따른 인식 정확도의 변화 및 그 추세선을 도시한 것이다. 붉은 점은 첫번째 layer만을 동결할 때, 푸른 점은 두번째 layer까지 동결하고, Fully Connected Layer에만 전이학습을 적용하였을 경우의 인식 정확도이다. 두 동결 깊이에 대하여, 사전훈련에 사용한 TMNIST 데이터베이스 대비 전이학습에 사용하는 MNIST 데이터의 비율을 0.5 %에서 10 %까지 늘려가며 인식 정확도를 추적하였다. 각 경우에 대한 추세선을 굵은 점선으로 표시하였다.

전이학습을 위하여 첫 번째 layer까지만 동결한 경우가 두 번째 layer까지 동결한 경우에 비하여 더 빨리 인식 정확도가 수렴하였다. 목표 인식 정확도를 90%로 기준하였을 때, 동결 깊이가 1인 경유는 TMNIST 데이터 대비 MNIST 데이터가 3% 이상이면 만족 가능하였으며, 7%이상일 경우 인식 정확도를 안정적으로 유지하였다. 반면 동결 깊이가 2일 경우, TMNIST 데이터 대비 MNIST 데이터가 5% 이상이어야 90% 인식 정확도에 도달 가능하였으며, 포화되지 않고 지속적으로 상승함이 관측되었다.

V. 결 론

본 연구는 로봇 비전에서 영상 인식을 비롯한 다양한 AI 분야에 활용되는 전이학습 그 자체에 대한 정량 평가 방법을 제안한다.

본 연구에서 활용한 MNIST 데이터베이스는 소량의 불균일하고 품질이 부족한 데이터로 현실의 제한된 데이터 수집 환경을 대변하며, TMNIST 데이터베이스는 대량의 균질한 데이터로 통제된 실험, 시뮬레이션 환경을 대변한다. 이를 기반으로 기준 네트워크의 사전훈련 진행 결과, 소량의 MNIST는 데이터 부족으로 학습이 수행되지 않았으며, TMNIST는 10 epoch 이내에 빠르고 정확하게 96.77 % 이상의 인식 정확도를 확보하였다. 전이학습 이전, TMNIST로 사전훈련된 기준네트워크는 MNIST에 대하여 75% 인식 정확도로 특징은 잡았으나 디지털 서체와 손글씨의 도메인 차이로 인식 정확도에 한계가 확인되었다. 이후, 첫 번째 layer를 동결하고 이후의 layer에 대하여 소량의 MNIST를 이용하여 전이학습을 적용한 결과, 91% 이상의 인식 정확도를 확보하였다.

이를 바탕으로 전이학습 그 자체에 대한 정량평가를 수행하기 위하여, 레이어 동결 깊이 및 사전학습 대비 전이학습 데이터의 비율을 조절하며 인식 정확도 변화를 추적하였다. 본 연구의 기준 네트워크에 대하여, 전이학습을 위한 동결 깊이는 첫번째 layer까지로 지정할 경우 보다 빠르게 인식 정확도가 높아 전이학습에 저합하였다. 이때, 사전훈련 데이터 대비 전이학습 데이터의 비율이 3% 이상이면 목표 인식 정확도인 90%에 안정적으로 도달 가능하다는 정량 기준의 확립이 가능하다.

따라서, 전이학습 파라미터에 따른 인식 정확도 변화 추세선은 전이학습 자체에 대한 조건을 적절히 적용하였는지에 대한 정량적 판단 기준으로 활용이 가능하다. 이를 바탕으로 로봇 비전용 영상 인식 AI 모델을 개발할 때, 전이학습을 위한 동결 깊이에 대한 기준을 정할 수 있으며, 또한 사용 가능한 사전훈련 데이터 대비 전이학습 데이터 비율에 따라 현재 전이학습 수준이 안정되었는지 정량적 예측과 검증에 활용할 수 있다.

그러나, 본 연구는 기준 네트워크로 단순 형태의 CNN을 사용하였으며, label이 균일한 2D 이미지 데이터베이스인 MNIST 및 TMNIST를 활용하였으므로 보다 다양한 네트워크 구조와 학습 데이터, 도메인에 대한 추가 검증이 필요하다. 향후 보다 다양한 네트워크(RNN, Auto Encoder, U-net 등)에 대한 적용, 다른 차원의 데이터 및 영역 (음성, spectrogram, 생체 신호 등)에 대한 확장을 고려할 수 있다. 일례로 양과 질 모두가 부족한 임상 데이터의 기반 의료 AI 분야에서 시뮬레이션 및 생체모사로봇 등을 이용한 사전학습과 전이학습의 적용으로 그 한계를 극복할 수 있을 것이다.

References

[1] D. Kragic and H. I. Christensen, “Advances in robot vision,” Rob. Auton. Syst., vol. 52, no. 1, pp. 1 - 3, Jul. 2005, doi: 10.1016/J.ROBOT.2005.03.007.

[2] N. Telagam, A. Thotakuri Assistant Professor, Tk. Assistant Professor, and M. Vucha Professor, “Survey on Robot Vision: Techniques, Tools and Methodologies,” Int. J. Appl. Eng. Res., vol. 12, pp. 6887 - 6896, 2017, Accessed: Apr. 01, 2024. [Online]. Available: http://www.ripublication.com

[3] W. Liang et al., “Advances, challenges and opportunities in creating data for trustworthy AI,” Nat. Mach. Intell. 2022 48, vol. 4, no. 8, pp. 669 - 677, Aug. 2022, doi: 10.1038/s42256-022-00516-1.

[4] S. Tayebi Arasteh et al., “Collaborative training of medical artificial intelligence models with non-uniform labels,” Sci. Reports 2023 131, vol. 13, no. 1, pp. 1 - 9, Apr. 2023, doi: 10.1038/s41598-023-33303-y.

[5] X. Zhou, T. Bai, Y. Gao, and Y. Han, “Vision-Based Robot Navigation through Combining Unsupervised Learning and Hierarchical Reinforcement Learning,” Sensors 2019, Vol. 19, Page 1576, vol. 19, no. 7, p. 1576, Apr. 2019, doi: 10.3390/S19071576.

[6] C. Kim, S. Yoon, M. Han, and M. Park, “Transfer Learning-based Generated Synthetic Images Identification Model,,” J. Converg. Cult. Technol., vol. 10, no. 2, pp. 465 - 470, 2024, doi: http://dx.doi.org/10.17703/JCCT.2024.10.2.465.

[7] N. Kwak and D. Kim, “Study On Masked Face Detection And Recognition using transfer learning,” Int. J. Adv. Cult. Technol., vol. 10, no. 1, pp. 294 - 301, 2022.

[8] “MNIST in CSV.”https://www.kaggle.com/dataset/oddrationale/mnist-in-csv (accessed Dec. 24, 2023).

[9] “TMNIST (Typeface MNIST).”https://www.kaggle.com/datasets/nimishmagre/tmnist-typeface-mnist (accessed Dec. 24, 2023).