

■ 연구과제 요약문

과제명(기간)	Targeted and unsupervised 불량 자동 분류 (2019.03.01. ~ 2020.02.29)
연구책임자	김 건 희 (gunhee@snu.ac.kr)
개요	- 연구개발 목표와 내용에 대해 소개 (서술식으로 작성해도 됨) 최근 다양한 분류 작업에서 심층 신경망(deep neural networks)은 우수한 성과를 거두었다. 하지만 이들 성과는 균형 학습 데이터(balanced dataset)에서 달성한 것이며 실제 산업 현장 제조과정에 있는 불균형 학습 데이터(imbalanced dataset)를 사용할 경우 테일 클래스(Tail)의 분류 정확도가 급격히 떨어진다. 본 연구에서는 이러한 데이터 불균형 편향을 교정하여 디스플레이 패널의 결함 분류 정확도를 대폭 향상시키는 새로운 알고리즘 Source-to-Target Adaptation Learning(STAL)을 개발하였다. 본 연구에서는 전이학습(transfer learning)을 이용하여 Source 클래스에서 Target 클래스로 정보를 전이하며 이를 통해 학습 데이터가 부족한 테일 클래스(Tail)의 과적합을 방지하고 분류 정확도를 대폭 향상시켰다. 개발한 알고리즘을 여러 벤치마크 데이터셋 뿐만 아니라 디스플레이 패널 학습 데이터에 적용하여 기존대비 향상된 성능을 달성하였다.
연구개발 결과	- 연구개발 결과를 가능한 한 비전문가가 이해할 수 있도록 설명 (서술식으로 작성해도 됨) 학계에서 가장 도전적이며 대용량인 불균형 데이터셋 Open-Long Tail Recognition(OLTR) 벤치마크와 실제 산업 현장에서 사용되는 디스플레이 패널 데이터셋에 대하여 개발한 알고리즘을 평가하였다. OLTR 벤치마크는 총 세 가지 데이터셋 ImageNet-LT, Places-LT, SUN-LT으로 구성되어 있는데, 개발한 알고리즘은 이들 모두에서 최고 성능을 달성하였다. 특히 Places-LT와 SUN-LT 데이터셋은 이전 연구들이 분류기로 사용한 ResNet-152 보다 파라미터가 40%밖에 안되는 ResNet-50을 사용하였음에도 불구하고 최고의 성능을 달성하였다. 산업 현장에 사용되는 디스플레이 패널 데이터셋에 대하여 총 세 가지 품목에 대하여 실험을 진행하였다. Xception 모델을 이용하여 학습을 하였을 시 기존 방법 대비 F1 점수와 분류 정확도 측면에서 세 가지 품목에 대하여 모두 매우 향상된 결과를 달성하였다.
활용분야 및 기대효과	- 연구개발 결과의 활용 분야와 기대 효과를 서술 (서술식으로 작성해도 됨) 산업적인 측면 딥러닝 기반 인공지능 기술은 학습을 위하여 많은 데이터를 필요로 한다. 현실적인 데이터 수집 과정 주에 불균형한 형태의 학습 데이터는 피할 수 없으며 이에 대한 해결이 더욱 중요해지고 있다. 본 연구에서 개발한 알고리즘은 이런 문제점을 극복하는데 널리 활용될 수 있다. 특히, 산업 현장에서 불량 분류 작업을 할 때 특정 분류 클래스에 대한 정보와 데이터가 부족한 경우, 제안된 알고리즘을 통해 전이 학습을 수행하면 부족한 정보를 가지고 있는 클래스의 분류 정확도 성능을 향상시킬 수 있다. 이론적인 측면 모델에 상관없이(model-agnostic) 전이학습을 진행하는 알고리즘을 개발하고 이를 불균형 데이터셋에 적용하여 크게 향상된 분류 성능을 얻었다. 개발한 알고리즘은 전통적인 방법들과 달리 모델을 수정할 필요가 없으며 파라미터 튜닝이 크게 필요하지 않은 장점을 가지고 있다. 또한 기존 방법대비 매우 적은 파라미터를 사용하면서도 높은 성능을 달성했다. 개발한 알고리즘은 클래스별 모델링을 할 필요가 없으며, 그래프 모델을 이용하여 클래스 별로 특화한 학습 알고리즘으로 발전시킬 수 있다.