

Deep Metric Learning을 활용한 합성곱 신경망 기반의 피부질환 분류 기술

(Skin Disease Classification Technique Based on Convolutional Neural Network Using Deep Metric Learning)

김강민*, 김판구**, 전찬준**

(Kang Min Kim, Pan-Koo Kim, Chanjun Chun)

요약

피부는 외부 오염으로부터 일차적으로 몸을 보호하는 역할을 한다. 피부병이 발생하게 되면 피부의 보호 기능이 저하되므로 신속한 진단과 치료가 필요하다. 최근 인공지능의 발달로 인해 여러 분야에 기술적용을 위한 연구가 이루어지고 있으며, 피부과에서도 인공지능을 활용해 오진율을 줄여 신속한 치료를 받을 수 있는 환경을 만들기 위한 연구가 진행되고 있다. 종래 연구들의 주된 흐름은 발생 빈도가 낮은 피부질환의 진단이었지만, 본 논문에서는 사람들에게 흔히 발생할 수 있고, 개인이 명확히 판별하기 힘든 티눈과 사마귀를 합성곱 신경망을 통해 분류하는 방법을 제안한다. 사용한 데이터셋은 3개의 클래스로 이루어져 있으며, 총 2,515장의 이미지를 가지고 있다. 학습 데이터 부족과 클래스 불균형 문제가 존재한다. 모델의 학습에는 deep metric 손실 함수와 교차 손실 함수를 이용해 각각 성능을 분석하였으며, 정밀도, 재현율, F1 점수, 정확도의 측면에서 비교한 결과 deep metric 손실 함수에서 더 우수한 성능을 보였다.

■ 중심어 : 합성곱 신경망 ; deep metric 손실 ; 클래스 불균형 ; 피부질환

Abstract

The skin is the body's first line of defense against external infection. When a skin disease strikes, the skin's protective role is compromised, necessitating quick diagnosis and treatment. Recently, as artificial intelligence has advanced, research for technical applications has been done in a variety of sectors, including dermatology, to reduce the rate of misdiagnosis and obtain quick treatment using artificial intelligence. Although previous studies have diagnosed skin diseases with low incidence, this paper proposes a method to classify common illnesses such as warts and corns using a convolutional neural network. The data set used consists of 3 classes and 2,515 images, but there is a problem of lack of training data and class imbalance. We analyzed the performance using a deep metric loss function and a cross-entropy loss function to train the model. When comparing that in terms of accuracy, recall, F1 score, and accuracy, the former performed better.

■ keywords : convolutional neural network ; deep metric loss ; class imbalance ; skin disease

견하는 즉시 신속한 치료가 필요하다[2].

최근 인공지능의 발전으로 인해 컴퓨터 비전(computer vision), 자연어 처리(natural language processing), 음악, 미술과 같은 다양한 분야에서 인공지능 기술이 쓰이고 있으며 [3-6], CT, MRI의 의료 이미지를 이용하여 질병을 판독하는 영상의학 분야뿐만 아니라[7-8], 피부 의학 분야에서도 오진 없이 빠르게 진단하고 치치하기 위해 인공지능 기술을 적용하고 있다[9-12]. Alam은 습진(eczema)의 색과 질감 특징 추출을 통해 환부를 찾고 SVM(Support Vector Machine)을 이용하여 학습시켜 분류하는 방식을 제안한다[9]. Sumithra은 SVM

* 학생회원, 조선대학교 컴퓨터공학과 학사과정

** 정회원, 조선대학교 컴퓨터공학과 교수

본 연구는 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019R1C1C1011597).

접수일자 : 2021년 08월 24일
수정일자 : 1차 2021년 11월 25일

제재 확정일 : 2021년 12월 28일
교신저자 : 전찬준 e-mail : cjchun@chosun.ac.kr

과 K-NN(k-nearest neighborhood,)을 이용해 흑색종(melanoma), 대상포진(Shingles), 지루성 각화증(Seborrheic keratosis) 등의 분류 방식을 제안했으며[10], 또 Seixas은 의사결정트리(decision tree)를 이용해서 하지 쾨양의 피부 병변을 감지하는 방법을 제안한다[11]. 후에 Zhang은 깊은 합성곱 신경망(Deep Convolutional Neural Network, DCNN)을 기반으로 기저세포암(basal cell carcinoma), 색소 세포성 모반(melanocytic nevus), 지루성 각화증 등을 분류하는 방법을 제안했다[12]. 이 종래의 연구들은 주로 피부암의 조기 선별, 병변의 양성(benign), 악성(malignant) 이진 분류, 흑색종의 여부와 같은 특정 주제에만 집중한다. 그러나 사람들이 피부과를 찾는 이유의 90% 이상은 악성 조직에 의한 문제가 아니므로[13], 기존 목표와는 다른, 사람들이 흔히 겪을 수 있는 피부병 판별에 관한 연구 또한 요구된다[14].

지도학습(supervised learning)에서는 최적화 방법과 학습 데이터의 질에 따라 성능이 달라진다. 이상적인 학습 데이터는 일반화되고 정제된 데이터가 균등하고 다양으로 준비되었을 때이며, 이러한 조건에서 학습된 모델은 일반적으로 더 좋은 성능을 보인다. 그러나 실제 학습 환경에서는 정제되지 않고 비식별된 데이터가 더 많은 이유로, 학습을 위한 최적의 데이터는 언제나 부족하다. 이로 인해 지도학습을 위한 데이터셋 구축 시, 데이터의 비식별, 불균형 등의 문제가 발생할 수 있다. 인공지능 기술의 발달에 따라 모델과 알고리즘 개발뿐만 아니라 데이터셋에서 발생하는 문제를 다루기 위한 연구들이 진행되었다[15-27].

지도학습에서는 비식별 데이터를 사용하기 위해 라벨링(labeling) 작업이 필요하다. 그러나 수집한 데이터 대부분이 비식별 데이터일 경우, 이 작업을 인력으로 처리하기에는 소모되는 시간과 비용이 너무나 비효율적이다. 이러한 자원 소모 최소화하며 비식별 데이터를 활용하려는 방안으로 비지도학습(unsupervised learning)과 준지도학습(semi-supervised learning)이 제안되었다. 비지도학습은 데이터에 목표값(label)이 없는 상태에서 k-평균 알고리즘(k-means clustering algorithm)과 같은 군집(clustering) 알고리즘을 이용하여 비슷한 데이터끼리 그룹을 형성시켜 새로운 데이터가 들어와도 어떤 군집에 속하는지 추론 가능한 방법이다[15]. 준지도학습은 지도학습과 비지도 학습을 결합한 방식으로 목표값이 있는 데이터와 없는 데이터를 같이 학습시켜 모든 데이터에 라벨링 할 필요성을 줄여주는 방법이다. 식별된 데이터가 적을 때 비식별된 데이터와 같이 사용해 분류기의 성능을 향상하는 목적으로 사용된다. 준지도학습에는 wrapper method, intrinsically semi-supervised 등과 같은 방법이 존재한다[16]. 위와 같은 학습방식을 사용하면 지도학습에서의 어려움인, 비식별된 데이터 문제의 해소 방안이 될 수 있다.

데이터셋에서 클래스 간 데이터 분포의 불균형함은 모델의 학습 과정에서 특정 클래스에 과적합(overfitting)을 일으켜 일 반화된 모델의 회득을 어렵게 한다[17]. 데이터가 균등하지 못한 경우를 해결하기 위해 제시된 방안이 데이터 샘플링(data sampling) 알고리즘이다[18]. 데이터 샘플링 알고리즘은 클래스별 데이터 분포를 맞춰주기 위한 알고리즘으로 언더 샘플링(under sampling)과 오버 샘플링(over sampling) 방식이 존재한다. 언더 샘플링은 분포가 편향된 클래스의 데이터를 감소시켜 희소 클래스와 데이터 수를 맞추도록 조절하는 방법이다. 언더 샘플링의 종류로는 RUS(Random Under Sampling) TL(Tomek Link), CNN(Condensed Nearest Neighbour CNN), ENN(Edited Nearest Neighbour) 등이 있다[19-22]. 오버 샘플링은 분포가 작은 클래스의 데이터 수를 분포가 큰 클래스의 데이터 수 만큼 증가시켜 주는 방법으로, ROS(Random Over Sampling), ADASYN(Adaptive Synthetic Sampling), SMOTE(Synthetic, Minority Oversampling Technique)와 같은 방식이 있다[19-24]. 이러한 방식을 이용하여 데이터 불균형 문제를 해결할 수 있지만, 언더 샘플링에선 정보 손실이, 오버 샘플링에서는 과적합 발생 위험이 커질 수 있다는 단점이 있다[25].

모델의 학습을 위한 데이터 부재에서 오는 위와 같은 문제를 해결하기 위한 또 다른 방법은 적은 수의 데이터로 모델의 성능을 높이는 방법이며, 이전 학습된 지식을 가지고 다른 지식에 적용을 꾀하는 전이학습[26]과 few-shot 데이터들을 가지고 유사도를 비교해 분류 방법들이 있다[27]. 본 논문에서는 few-shot learning의 일종으로 데이터의 유사도 측정하는 metric learning을 사용한다[28].

metric learning은 유clidean 거리(euclidean metric)와 마할라노비스 거리(mahalanobis distance)와 같은 distance metric을 기반으로 기존에 데이터의 특징(feature)으로는 분류가 쉽지 않았던 데이터에 대해 클래스별로 구별할 수 있도록 거리 공간(metric space) 학습을 통해 데이터 간의 유사도를 측정하는 방법이다. metric learning의 목표는 입력 데이터의 특징 공간(feature space)을 거리 계산이 쉬운 임베딩 공간(embedding space)으로 연결(mapping)하는 임베딩 함수 f 의 학습이며, 학습방식은 다음과 같다. 임베딩 간의 거리가 D , x 와 y 를 입력 데이터, 임베딩 함수를 f 라고 할 때 x , y 가 같은 클래스라면 임베딩 거리값을 작게, 다른 클래스라면 임베딩 거리값을 크게 만들도록 임베딩 함수 f 를 학습한다[29]. metric learning은 일반적으로 선형 투영(linear projection) 방식을 사용한다. 그러나 이러한 방식은 비선형적(non-linear) 성질을 갖는 현실 문제를 해결하는 데 한계가 있다.

이를 극복하기 위해 등장한 것이 deep metric learning이다. deep metric learning은 깊은 신경망(deep neural network)

에 metric learning을 적용하는 방식을 의미하며, 신경망의 비선형 구조를 갖는 활성화 함수(activation function)를 사용하여 현실의 비선형적 문제를 해결하기 위해 제안되었다[29]. 본 논문에서는 질병 이미지를 데이터로 사용하기에 특징 추출을 위한 깊은 신경망으로 이미지 처리에 활용해 쓰이는 합성곱 신경망을 백본(backbone)으로 사용하였으며, 합성곱 신경망에 deep metric learning을 적용하여 질병의 분류를 시도한다.

deep metric learning은 데이터에서 레이블로 직접 매핑하여 학습하는 것보다 강건성과 안정성이 장점이 있는 것이 보고되었으며, 얼굴 인식, 사람 재인식(person re-identification), visual tracking, 3D 형상 검색(3D shape retrieval)과 같은 분야에 적용되고 있다[30-33]. 의료 분야에서도 방사선 사진을 데이터로 이용해 질병의 진단과 예측에 사용하는 등의 연구가 활발하게 이루어지고 있다[34-35].

본 논문에서는 피부질환에 전문성이 없는 사람들이 사마귀와 터눈이 비슷한 외형으로 인해 잘못된 판단을 내리지 않도록 돋는 피부질환 분류 모델을 제안한다. 본 논문에서는 부족한 데이터로 인해 많은 데이터를 기반으로 특징을 학습하는 지도학습으로는 성능을 기대하기 어려운 상황이다. 이러한 조건에서 데이터 증강이나 샘플링 알고리즘의 사용은 정보 손실, 과적합의 위험을 더욱 높일 수 있으므로 few-shot learning 방식인 deep metric learning을 사용하여 데이터의 특징보다는 유사도를 기반으로 분류하는 모델을 학습시킨다. 이러한 모델은 적은 데이터의 양으로도 비슷한 외형으로 인해 구분이 어려운 데이터를 분류하는 데 강점을 가질 수 있으며, 이로 인해 본 논문이 다루는 문제를 해결하기에 적합하다. 위와 같은 방식을 통해 관련 연구가 적은 사마귀와 터눈의 분류 방법을 제안함과 동시에 사용한 deep metric learning 손실 함수가 분류 성능에 미치는 영향에 대해 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 모델의 학습에 사용한 deep metric learning 손실 함수들을 소개한다. 3장에서는 본 논문에서 사용한 합성곱 신경망 모델의 구조와 데이터셋을 설명한다. 4장에서는 모델과 손실 함수를 통해 도출된 결과를 분석한다. 마지막 5장에서는 본 논문의 결론과 후속 연구에 대해 논의한다.

II. Deep Metric Learning 손실 함수

1. Contrastive 손실 함수

Contrastive 손실은 삼 네트워크(siamese network)에 쓰이는 손실 함수이다. 삼 네트워크는 닮은 두 모델이 입력 데이터의 유사성을 학습하기 위해 같은 가중치를 공유하는 신경망이다. Contrastive 손실을 이용한 Contrastive 임베딩은

$\{(x_i, x_j, y_{ij})\}$ 와 같이 쌍으로 이루어진 데이터를 학습한다. x_i, x_j 는 입력 데이터를 나타내며, y_{ij} 는 이 두 데이터의 클래스가 같은 쌍(positive pair)인지 다른 쌍(negative pair)인지를 나타낸다. 같은 클래스일 경우 1, 다른 클래스의 경우 0의 값을 갖는다. Contrastive 손실은 두 입력 데이터가 같은 클래스라면 입력 데이터 간의 거리는 가깝게 임베딩하며, 다른 클래스라면 α 만큼의 차이(margin)를 갖도록 임베딩 한다. 손실 함수의 수식은 아래의 수식 1을 따른다[36].

$$L = \frac{1}{2} \sum_{i,j=1}^n y_{i,j} D_{i,j}^2 + (1 - y_{i,j}) [\alpha - D_{i,j}]_+^2 \quad (1)$$

$D_{i,j}$ 는 $\|f(x_i) - f(x_j)\|_2$ 의미하며, 이때 $f(\cdot)$ 는 모델을 통해 거쳐서 나오는 이미지의 임베딩 특정값을 나타낸다. $[\cdot]_+$ 은 $\max(0, \cdot)$ 의 hinge 손실 함수 연산을 나타낸다.

2. Triplet 손실 함수

Triplet 손실은 $\{(x_a^{(i)}, x_p^{(i)}, x_n^{(i)})\}$ 와 같이 삼중항으로 이루어진 훈련 데이터를 통해 Triplet 임베딩을 수행한다. $(x_a^{(i)}, x_p^{(i)})$ 는 서로 같은 클래스의 데이터이며 $(x_a^{(i)}, x_n^{(i)})$ 는 다른 클래스의 데이터이다. $(x_a^{(i)}, x_p^{(i)})$ 의 임베딩 거리는 가깝게 $(x_a^{(i)}, x_n^{(i)})$ 의 거리는 $(x_a^{(i)}, x_p^{(i)})$ 에 차이 α 를 더한 거리보다 멀리 배치하도록 학습하는 것을 목적으로 한다. 수식으로 표현하면 수식 2

$$\|f(x_a^{(i)}) - f(x_p^{(i)})\|_2^2 + \alpha < \|f(x_a^{(i)}) - f(x_n^{(i)})\|_2^2 \quad (2)$$

$$L = \frac{3}{2m} \sum_{i=1}^m [\|f(x_a^{(i)}) - f(x_p^{(i)})\|_2^2 - \|f(x_a^{(i)}) - f(x_n^{(i)})\|_2^2 + \alpha]_+ \quad (3)$$

와 같으며, 이를 기반으로 Triplet 손실은 수식 3과 같이 표현할 수 있다[30].

3. Lifted Structured 손실 함수

Lifted Structured 손실은 학습 배치(batch) 데이터를 최대한 활용하며 모든 같은 클래스의 쌍(positive pair)이 다른 클래스 쌍(negative pair)과 거리를 비교하는 방법이다. 배치 내에서 데이터 쌍별 거리(pairwise distance) 벡터(vector)를 올려(lifting) 완전 연결 밀도행렬(fully connected dense matrix)로 변형시킨 후 배치 내 데이터들의 모든 쌍 관계를 계산한다. 이와 같은 방법으로 데이터를 충분히 활용하며 거리를 계산할 수 있다. Lifted Structured 손실 함수는 수식 4와 같다[37].

$$L = \frac{1}{2|P|} \sum_{(i,j) \in P} \left[\log \left(\sum_{(i,k) \in N} e^{\alpha - D_{i,k}} + \sum_{(j,l) \in N} e^{\alpha - D_{j,l}} \right) + D_{i,j} \right]_+^2 \quad (4)$$



그림 1. 구축한 피부질환 이미지 데이터

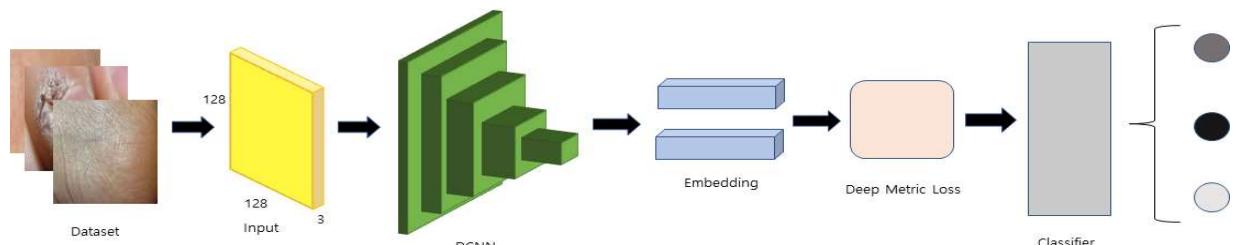


그림 2. deep metric learning 손실을 이용한 분류 모델

이때 P는 클래스 같은 짹을 나타내며 N은 클래스가 다른 짹을 나타낸다.

표 1. 이미지 데이터 세트 구성

피부질환	맨살 (Normal skin)	티눈 (Corns)	사마귀 (Warts)	합계
Train (80%)	808	350	855	2013
Test (20%)	202	87	213	502
소계 (%)	1,010 (40.2)	437 (17.4)	1,068 (42.4)	2,515

III. 합성곱 신경망을 활용한 피부질환 분류

본 논문에서는 합성곱 신경망을 이용해 비슷한 외형을 가진 사마귀와 티눈의 분류 방법을 제안한다[38]. 피부 질병 분류에

쓰이는 데이터는 HAM1000, DermIS, DermNZ 등에서 얻을 수 있지만[39], 본 논문에서 사용하는 사마귀와 티눈에 대한 데이터는 존재하지 않거나 수가 매우 적었기에, 구글을 통해 환자들이 환부를 직접 찍은 이미지들을 수집하여 모델 학습을 위한 추가 데이터로 사용하였다. 이러한 데이터들로 학습된 모델은 실제 개인들이 일상 중 질병으로 의심되는 부위를 찍어 확인할 시에 높은 예측률을 보일 수 있을 것으로 기대된다. 수집한 데이터는 크기가 모두 다르고 병변 부위와 비병변 부위가 같이 존재하는 상태이다. 우리는 모델이 병변에만 더욱 집중하여 학습 되기를 원하기에 비병변 부위를 제거하고 병변 부위만 남기도록 데이터 전처리를 수행하였다.

표 1은 합성곱 신경망 모델을 학습하기 위하여 활용된 데이터 셋의 구성이며, 그림 1은 이미지 데이터의 예시이다. 클래스는 사마귀(Warts), 티눈(Corns), 맨살(Normal Skin) 세 가지로 구분된다. 전체 2,515장의 이미지 데이터를 가지고 있으며, 이 중 사마귀는 전체 비율의 42.4%를 차지하고 맨살은 40.2%를 차지한다. 희소 클래스인 티눈은 17.4%를 차지한다. 각 클래스 데이터의 80%는 학습용으로 사용하며, 20%는 평가데이터로 사용한다. 구축한 데이터셋을 보면 학습을 위한 이미지 데이터의 부족과 각 클래스 간의 불균형 문제가 존재하는 것을 확인할

수 있다. 이러한 조건에서 [12]와 같은 합성곱 신경망 분류 방식으로 학습을 시킬 시 일반화된 모델을 얻을 수 없다. 위와 같은 문제에서 최적의 학습모델을 얻기 위해 퓨샷 데이터도 잘 구분

표 2. 합성곱 신경망 모델과 손실 함수에 따른 분류 성능 비교

(단위 : %)	InceptionV3				ResNet50				MobileNet			
	정밀도	재현율	F1 점수	정확도	정밀도	재현율	F1 점수	정확도	정밀도	재현율	F1 점수	정확도
CE	85.57	85.88	85.61	85.68	54.12	64.81	58.42	64.81	93.23	92.84	92.97	92.84
C	93.44	93.43	93.40	93.44	86.52	85.68	86.05	85.68	95.42	95.22	95.26	95.22
TP	90.89	90.65	90.73	90.66	85.89	84.29	84.55	85.68	95.05	95.02	95.02	95.02
LS	89.92	89.86	89.73	89.86	92.26	92.24	92.17	92.25	97.43	97.41	97.42	97.41

할 수 있는 deep metric learning으로 데이터들의 임베딩 학습을 시킨 후 분류를 진행한다. 모델 구조의 백본(backbone)으로는 Inception, ResNet, MobileNet을 사용하며[40-42], 학습을 위한 deep metric learning 손실 함수는 Contrastive 손실(C), Triplet 손실(TP), Lifted Structured 손실(LS)을 사용한다. deep metric learning 손실 함수를 사용하지 않고 분류학습을 진행했을 시 어떤 결과를 보이는지 확인하기 위해 교차 엔트로피(Cross-Entropy Loss, CE) 손실을 이용한 학습 결과를 비교한다.

그림 2는 deep metric learning을 적용한 합성곱 신경망 모델의 전체도이다. 전처리를 통해 병변 부위만 추려낸 데이터들의 크기는 128*128이었으며, 약간의 오차가 있는 데이터들은 축소 또는 패딩 되어 128*128 크기로 변환 후 모델의 입력으로 들어간다. 깊은 신경망을 거친 후 이미지의 특징값은 임베딩 값으로 변환된다. 그 후 deep metric learning 손실 함수를 통해 입력 데이터의 짝에 따라 임베딩 거리값이 유지될 수 있게 임베딩 함수를 학습한다. 이후 완전 연결 계층(fully-connected layer)을 이어 분류하는 과정을 거친다.

분류기의 학습에 최적화(optimizer)는 Adam을 이용했으며, 활성화 함수(activation function)로는 relu, 학습률(learning rate)은 deep metric 손실 함수에서 5×10^{-4} , 분류학습에서 6×10^{-4} 으로 설정하였다. 조기 종료(early stopping) 알고리즘은 학습 loss 값을 확인하여 3회 동안 감소가 없을 시 종료하도록 설정하였다.

IV. 성능평가

표 2는 각 모델에서 손실 함수를 이용한 분류 성능을 나타낸다. 평가지표는 정밀도(precision), 재현율(recall), F1 점수(F1 Score), 정확도(accuracy)를 사용한다.

InceptionV3 모델에서는 Contrastive 손실, Triplet 손실, Lifted Structured 손실 모두 교차 엔트로피 손실보다 성능향상을 보였다. 이 중에서 가장 효과가 큰 것은 Contrastive 손실

표 3. 합성곱 신경망 모델의 파라미터 개수 비교

Models	Total params
ResNet	23,587,712
InceptionV3	21,802,784
MobileNet	3,228,864

이며, 교차 엔트로피 손실과 비교하여 정밀도에서 7.9 %, 재현율에서 7.6%, F1 점수에서 7.8%, 7.7% 정확도가 향상되었다.

ResNet 모델에서는 다른 모델보다 교차 엔트로피 손실의 학습 결과가 낮은 것뿐만 아니라 정밀도, 재현율, F1 점수, 정확도의 차이가 큰 것이 확인되었다. 혼합행렬(confusion)로 확인한 결과 사마귀와 맨살 클래스에 대한 학습만 이루어지고 희소 클래스인 티눈에 대한 정답은 전혀 맞히지 못하는 결과를 보였다. 다른 모델보다 ResNet에서 불균형 데이터셋으로 인한 특정 클래스 과적합 정도가 높았다. ResNet 모델에서 가장 큰 성능향상을 보인 것은 Lifted Structured 손실이었으며, 교차 엔트로피 손실보다 정밀도에서 38.1%, 재현율과 정확도에서 27%, F1 점수에서 33% 성능향상이 있었다.

MobileNet 모델에서는 Contrastive 손실, Triplet 손실, Lifted Structured 손실 모두 교차 엔트로피 손실보다 성능이 높았다. 이 중에서 Lifted Structured 손실 적용 효과가 가장 커졌으며, 교차 엔트로피 손실과 비교하였을 때, 정밀도에서 4.2%, 재현율에서 4.6%, F1 점수에서 4.5%, 정확도에서 4.1%의 향상을 보였으며, 정확도가 97.41%로 세 모델 중 높은 결과값을 보였다. 모델 성능에 대해 분석한 결과 다른 모델보다 MobileNet에서 손실 함수와 관계없이 다른 모델보다 좋은 성능을 보였다. 표 3은 모델의 파라미터(parameter) 개수를 보여

표 4. MobileNet 모델과 교차 엔트로피 손실을 활용한 학습 평가 혼합행렬

Predicted label \ True label	사마귀 (Warts)	티눈 (Corns)	맨살 (Normal Skin)
사마귀 (Warts)	0.9273	0.0591	0.0136
티눈 (Corns)	0.1481	0.8519	0.0000
맨살 (Normal Skin)	0.0050	0.0248	0.9703

표 5. MobileNet 모델과 Lifted Structured 손실을 활용한 학습 평가 혼합행렬

Predicted label \ True label	사마귀 (Warts)	티눈 (Corns)	맨살 (Normal Skin)
사마귀 (Warts)	0.9714	0.0238	0.0048
티눈 (Corns)	0.0380	0.9494	0.0127
맨살 (Normal Skin)	0.0093	0.0047	0.9860

준다. InceptionV3 와 ResNet 모델의 파라미터의 수가 데이터셋 보다 많이 웃돌아 학습 데이터에서 과적합 문제가 발생했다. 이번 데이터셋과 같은 조건이면 모델 파라미터가 상대적으로 적은 MobileNet이 더 적합함을 확인하였다.

다른 모델과 비교해 성능이 가장 잘 나왔던 MobileNet에서 유사도를 사용하지 않는 교차 엔트로피 손실과 가장 준수한 성능을 보였던 유사도 기반 손실 함수인 Lifted Structured 손실의 결과를 비교하였다.

표 4-5는 교차 엔트로피와 Lifted Structured 손실학습 결과를 혼합행렬로 시각화한 결과이다. 교차 엔트로피를 사용한 결과에서는 상대적으로 데이터가 많은 두 클래스는 90%가 넘는 정확도를 보이는 대에 비해 티눈에서는 데이터 부족으로 인해 85%의 정확도를 보였다. Lifted Structured 손실을 이용한 결과에서는 티눈 클래스에 대해서도 비교적 균등한 학습이 이루어져 교차 엔트로피의 결과에서 9.8% 성능이 향상된 94% 정확

도를 보인 것을 확인할 수 있다. 사마귀와 맨살 클래스에 대해 서도 교차 엔트로피의 결과에서 각각 4.4%, 1.6% 향상된 성능을 보였다.

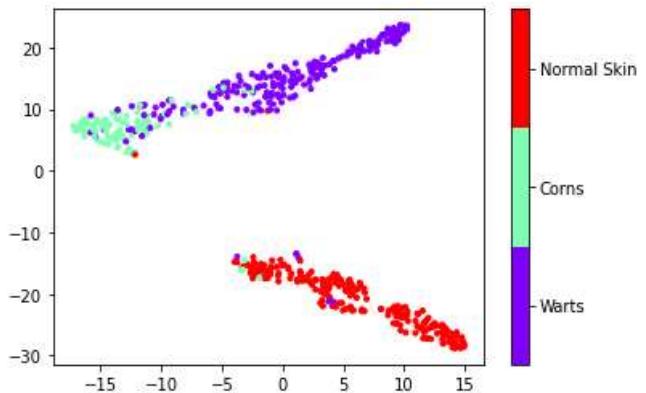


그림 3. t-SNE로 시각화한 MobileNet 모델과 교차 엔트로피 손실을 활용한 학습 평가 결과

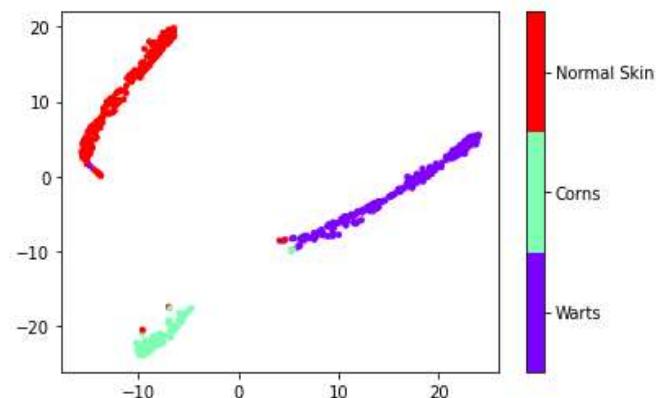


그림 4. t-SNE로 시각화한 MobileNet 모델과 Lifted Structured 손실을 활용한 학습 평가 결과

그림 3-4는 t-분포 확률적 임베딩(t-Stochastic Neighbor Embedding, t-SNE)을 이용하여 MobileNet에서의 교차 엔트로피와 Lifted Structured 손실학습의 평가 결과를 시각화한 것이다[43]. 붉은색은 질병이 없는 정상적인 피부를 나타내며, 청록색은 티눈을, 푸른색은 사마귀를 나타낸다.

그림 3은 교차 엔트로피를 사용한 결과이다. 티눈과 사마귀의 외형적으로 유사한 특징으로 인해 두 군집의 거리가 밀접해 있는 것을 볼 수 있으며, 그림 3과 표 4를 보면 사마귀와 티눈을 분류하는 데에 맨살과 비교하여 잘못 분류하는 경우가 많음을 확인할 수 있다. 전체적으로 예측 결판값이 다른 군집에 섞여 있는 경우가 많아 교차 엔트로피를 사용한 결과는 분류학습이 잘 이루어졌다고 보기 어렵다.

그림 4는 Lifted Structured 손실을 사용한 결과이며, 그림 3과 비교해 각 군집의 밀도가 높고, 군집 사이의 거리도 뚜렷하

게 분리되어있는 것을 확인할 수 있다. Lifted Structured를 사용한 결과에서도 각각 클래스의 예측 결과값들이 다른 군집에 섞여 있는 결과를 보이지만, 그림 3보다 그 수가 확연히 적어 분류학습이 비교적 잘 이루어졌음이 보인다.

V. 결 론

본 논문에서는 합성곱 신경망에 metric learning을 적용한 deep metric learning 방식으로 사마귀와 티눈을 분류하는 방법을 제안하였다. 구축된 데이터셋이 충분하지 않은 관계로 데이터 증강, 샘플링 알고리즘으로 과적합의 위험을 높이는 방식 보단 few-shot learning에 사용되는 metric learning 방식을 선택하였다. 실험 결과 본 논문의 조건과 같이 데이터셋이 부족한 상태에서는 MobileNet을 백본으로 선택했을 시 모든 손실 함수에서 가장 좋은 성능을 보였으며, MobileNet에서 Lifted Structured 손실을 사용했을 때 정밀도, 재현율, F1 점수, 정확도에서 가장 준수한 성능을 보였다. 유사도 기반 방식이 아닌 교차 엔트로피 손실 방식과 비교하였을 때 각 평가지표의 성능과 향상도 측면에서 정밀도는 97.4%(4.2%), 재현율은 97.41%(4.6%), F1 점수는 97.4%(4.5%), 정확도 97.4%(4.1%)의 성능향상을 보였다. deep metric learning으로 인한 성능향상은 다음과 같은 이유로 보인다. 이미지 특징 추출을 통한 분류 모델은 특징을 학습하기 위한 데이터가 충분하지 못할 때 유사한 이미지에 대한 차이를 파악하지 못해 특징이 비슷하게 나와 모델이 분류에 어려움을 겪을 수 있다. deep metric learning은 데이터들의 특징값을 임베딩 값으로 변환 후 데이터끼리 임베딩 값의 거리 비교를 수행한다. 데이터 임베딩 거리 값이 데이터 쌍에 따라 같으면 가깝게, 다르면 멀어지도록 가중치를 업데이트하는데 유사해 보이는 데이터라도 같음, 다름이라는 차이를 파악할 수 있는 기준이 있어 이미지 분류에 더욱 효과적인 것으로 보인다.

본 실험을 통해 deep metric learning 방식이 사마귀와 티눈 분류 성능에 미치는 영향을 분석하였으며, 이외의 피부질환 분류에 대해서도 긍정적인 효과를 보일 수 있을 것으로 기대된다.

향후 연구에서는 다른 피부 질병에 대한 클래스가 늘었을 때 deep metric learning이 어떤 성능을 보이는지 뿐만 아니라 이미지와 더불어 환자의 의료기록과 같은 메타데이터가 결합하였을 때 어떤 성능향상을 보이는지에 대한 실험과 이와 관련 데이터의 수집이 필요하겠다.

감사의 글

본 연구는 2021년도 정부(과학기술정보통신부)의 지원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019R1C1C1011597).

REFERENCES

- [1] M. A. M. Almeida, and I. A. X. Santos, "Classification models for skin tumor detection using texture analysis in medical images," *Journal of Imaging*, vol. 6, no. 6, Jun., 2020.
- [2] V. Ki, and C. Rotstein, "Bacterial skin and soft tissue infections in adults: A review of their epidemiology, pathogenesis, diagnosis, treatment and site of care," *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 19, no. 2, pp. 173–184, Mar., 2008.
- [3] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, Jul., 2017.
- [4] D. Castelvecchi, "Deep learning boosts google translate tool," *Nature News*, Sep., 2016.
- [5] M. Blaauw, and J. Bonada, "A neural parametric singing synthesizer modeling timbre and expression from natural songs," *Applied Sciences*, vol. 7, no. 12, Dec., 2017.
- [6] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423, Jul., 2016.
- [7] J. Cheng, D. Ni, Y. Chou, J. Qin, C. Tiu, Y. Chang, C. Huang, D. Shen, and C. Chen, "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in CT scans," *Scientific Reports*, Apr., 2016.
- [8] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, May, 2016.
- [9] M. N. Alam, T. T. K. Munia, K. Tavakolian, F. Vasefi, N. MacKinnon, and R. Fazel-Rezai, "Automatic detection and severity measurement of eczema using image processing," *Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1365–1368, Aug., 2016.
- [10] R. Sumithra, M. Suhil, and D. S. Guru, "Segmentation and classification of skin lesions for disease diagnosis," *Procedia Computer Science*, vol. 45 pp. 76–85, Mar., 2015.

- [11] J. L. Seixas, and R. G. Mantovani, "Decision trees for the detection of skin lesion patterns in lower limbs ulcers," *Proceedings of the International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 677-681, Dec., 2016.
- [12] X. Zhang, S. Wang, J. Liu, and C. Tao, "Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge," *BMC Medical Informatics and Decision Making*, vol. 18, pp. 69-76, Jul., 2018.
- [13] E. N. Wilmer, C. J. Gustafson, C. S. Ahn, S. A. Davis, S. R. Feldman, and W. W. Huang, "Most common dermatologic conditions encountered by dermatologists and nondermatologists," *Cutis*, vol. 94, no. 6, pp. 285-292, Dec., 2014.
- [14] Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, K. Kanada, G. de O. Marinho, J. Gallegos, S. Gabriele, V. Gupta, N. Singh, V. Natarajan, R. H. Wellenhof, G. S. Corrado, L. H. Peng, D. R. Webster, D. Ai, S. J. Huang, R. C. Dunn, and D. Coz, "A deep learning system for differential diagnosis of skin diseases," *Nature Medicine* vol. 26, pp. 900-908, May, 2020.
- [15] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 132-149, Sep., 2018.
- [16] J. E. Engelen, and H. H. Hoos, "A survey on semi-supervised learning", *Machine Learning*, vol. 15, pp. 373-440, Nov., 2019.
- [17] R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, "Data mining with imbalanced class distributions: concepts and methods," *Proceedings of the 4th Indian International Conference on Artificial Intelligence(IICAI-09)*, pp. 359-376, Dec., 2009.
- [18] E. Rendin, R. Alejo, C. Castorena, E. J. Isidro-Ortega, and E. E. Granda-Gutierrez, "Data sampling methods to deal with the big data multi-class imbalance problem," *Applied science*, vol. 10, Feb., 2020.
- [19] N. Japkowicz, and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429-449, Nov., 2002.
- [20] I. Tomek, "Two modifications of CNN," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, no. 11, pp. 769-772, 1976.
- [21] P. Hart, "The condensed nearest neighbour rule," *IEEE Transactions on Information Theory*, pp. 515-516, 1968.
- [22] D. Wilson, "Asymptotic properties of nearest neighbor rules using edited data," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-2, pp. 408-420, Jul., 1972.
- [23] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 1322-1328, Jun., 2008.
- [24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, Jun., 2002.
- [25] C. Chang, M. Hsu, E. X. Esposito, and Y. J. Tseng, "Oversampling to overcome overfitting: Exploring the relationship between data set composition, molecular descriptors, and predictive modeling methods," *Journal of Chemical Information and Modeling*, vol. 53, no. 4, pp. 958-971, Mar., 2013.
- [26] S. J. Pan, and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct., 2010.
- [27] J. Kim, T. Kim, S. Kim, and C. D. Yoo "Edge-labeling graph neural network for few-shot learning," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11-20, Jun., 2019.
- [28] W. Jiang, K. Huang, J. Geng and X. Deng, "Multi-Scale Metric Learning for Few-Shot Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 1091-1102, Mar., 2021.
- [29] M. Kaya, and H. S. bilge, "Deep metric learning: A survey," *Symmetry*, vol. 11, no. 9, Aug., 2019.
- [30] F. Schroff, D. Kalenichenko, and J. Philbin. "Facenet: A unified embedding for face recognition and clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815-823, Jun., 2015.
- [31] A. Hermans, L. Beyer, and B. Leibe, "In defense

- of the triplet loss for person re-identification,” *arXiv Preprint arXiv:1703.07737*, 2017.
- [32] J. HU, J. Lu, and Y-P. Tan, “Deep metric learning for visual tracking,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 2056–2068, Nov., 2016.
- [33] J. Xie, G. Dai, F. Zhu, L. Shao and Y. Fang, “Deep Nonlinear Metric Learning for 3-D Shape Retrieval,” *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 412–422, Jan., 2018.
- [34] M. Annarumma and G. Montana, “Deep metric learning for multi-labelled radiographs,” *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pp. 34–37, Apr., 2018.
- [35] A. Zhong, X. Li, D. Wu, H. Ren, K. Kim, Y. Kim, V. Buch, N. Neumark, B. Bizzo, W. Y. Tak, S. Y. Park, Y. R. Lee, M. K. Kang, J. G. Park, B. S. Kim, W. J. Chung, N. Guo, I. Dayan, M. K. Kalra and Q. Li, “Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19,” *Medical Image Analysis*, vol. 70, May, 2021.
- [36] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun., 2006.
- [37] H. O. Song, Y. Xiang, S. Jegelka, and S. Savarese “Deep metric learning via lifted structured feature embedding,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4004–4012, Jul., 2016.
- [38] 이정덕, “물리적 인자에 의한 피부손상,” *Journal of Korean Medical Association*. vol. 62, pp. 197–201, Apr., 2019.
- [39] N. Hameed, A M. Shabut, and M. A. Hossain, “Multi-class skin diseases classification using deep convolutional neural network and support vector machine,” *Proceedings of the International Conference on Software, Knowledge Information, Industrial Management and Applications (SKIMA)*, pp. 1–7, Dec., 2018.
- [40] J. Velasco, C. Pascion, J. W. Alberio, J. Apuang, J. Cruz, M. A. Gomez, B. Molina, L. Tuala, A. Thio-ac, R. and J. Jordà, “A smartphone-based skin disease classification using mobilenet CNN,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, pp. 2632–2637, Oct., 2019.
- [41] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, pp. 115–118, Jan., 2017.
- [42] A. Budhiman, S. Suyanto, and A. Arifianto, “Melanoma cancer classification using resnet with data augmentation,” *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 17–20, Dec., 2019.
- [43] L. V. D. Maaten, and G. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, Nov., 2008

저자 소개



김강민(학생회원)

2017년 ~ 현재 조선대학교 컴퓨터공학과 학사과정.

<주관심분야 : 딥러닝, 인공지능, 자연어처리>



김판구(정회원)

1988년 조선대학교 컴퓨터공학과 학사 졸업.
1990년 서울대학교 컴퓨터공학과 석사 졸업.
1994년 서울대학교 컴퓨터공학과 박사 졸업.

1994 ~ 현재 조선대학교 컴퓨터공학과 교수

<주관심분야 : 지능형정보처리, 시멘틱 웹, 온톨로지, 자연어처리, 데이터 마이닝 등>



전찬준(정회원)

2009년 한국기술교육대학교 정보기술 공학부 학사 졸업.
2011년 광주과학기술원 정보기전공학부 석사 졸업.
2017년 광주과학기술원 전기전자컴퓨터공학부 박사 졸업.

2021 ~ 현재 조선대학교 컴퓨터공학과 조교수

<주관심분야 : 음성 및 오디오 신호처리, 음원위치추정, 딥러닝, 인공지능>