Journal of KIIT. Vol. 19, No. 2, pp. 1-10, Feb. 28, 2021. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2021.19.2.1

클래스 불균형 상황에서 VAE를 활용한 패션 스타일 분류 성능 향상

박 종 혁*

Improving Fashion Style Classification Accuracy using VAE in Class Imbalance Problem

Jonghyuk Park*

요 약

최근 들어 인공지능이 활발하게 연구됨에 따라 이를 활용한 서비스 및 시스템이 여러 분야에서 제안되고 있다. 패션 또한 그 분야 중 하나로 패션 아이템 분류, 검출 등을 활용한 서비스가 개발되고 있다. 본 연구에서 다루고 있는 패션 스타일 분류 문제의 경우, 데이터 수집 과정에서 스타일의 유행에 의한 클래스 불균형 문제가 발생한다. 데이터 수가 적은 클래스에 대해서 낮은 분류 성능을 유발하는 클래스 불균형 문제에 대처하기 위하여 본 연구에서는 VAE(Variational AutoEncoder)를 활용한다. 패션 이미지로부터 학습된 VAE 잠재 변수의 확률 분포를 통해, 제안 모델은 데이터 수가 적은 클래스에 대해서는 더 많은 잠재 변수를 샘플링한다. 한 이미지를 표현하는 다양한 잠재 변수로부터 분류기가 학습되기 때문에 데이터 오버샘플링으로 학습할 때 발생하는 과적합을 피할 수 있다. 실험 결과, 제안 모델은 데이터 오버샘플링을 통해 학습된 모델과 비교 시 패션 스타일 분류 성능을 향상시키는 것을 확인할 수 있었다.

Abstract

As artificial intelligence is actively researched in recent years, services and systems utilizing it have been proposed in various fields. In the fashion domain, which is one of those fields, services that employ fashion item classification method or detection method are introduced. In the case of fashion style classification addressed in this study, a class imbalance problem occurs due to the fashion trends in the data collection process. Therefore, we adopt VAE (Variational AutoEncoder) to cope with the class imbalance problem that causes low classification performance for the classes with few data. Through the probability distribution of VAE's latent variable, the proposed model samples more latent variables for the classes with few data. Since the classifier is learned from these various latent variables representing a single image, overfitting that occurs when learning with an oversampling method is avoided. As a result of experiments, the proposed model improved performances compared with the model trained with an oversampling method.

Keywords

image classification, fashion style, variational autoencoder, class imbalance problem, deep learning

* 서울대학교 산업공학과 박사과정(교신저자)
- ORCID: https://orcid.org/0000-0003-4283-1155

· Received: Nov. 13, 2020, Revised: Jan. 13, 2021, Accepted: Jan. 16, 2021

· Corresponding Author: Jonghyuk Park
Dept. of Industrial Engineering, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea
Tel.: +82-2-880-7631, Email: chico2121@snu.ac.kr

Ⅰ. 서 론

최근 인공지능 연구의 발전으로 인하여 산업 각 분야에 이를 활용한 서비스 혹은 정보 시스템이 제안되고 있다[1]-[3]. 패션 업계 또한 그런 산업 분야 중 하나로, CNN(Convolutional Neural Network)[4] 기반 모델을 적용함으로써 여러 문제의 성능을 올리고 있다[5]-[10].

패션 관련 인공지능 연구는 이미지 분류, 의상 아이템 검출, 비슷한 패션 추출, 의상의 랜드마크 검출 등의 분야에서 이루어지고 있다. 그 중 패션 이미지 분류 문제에 관련된 연구는 의상 종류 및 특징 분류에 집중되어 있는데, 이미지 속에 등장하는 옷의 모양이나 패턴 같은 외형적인 요소를 인식하는 것이 성능 향상에 중요한 부분을 차지한다. 이에 비해 패션 스타일에 따른 이미지 분류 문제는 옷의 외형뿐만 아니라 인간의 인지와 밀접한 연관이 있는 문제이다. 특히 같은 스타일을 표현하는 이미지가 다른 패션 분류 문제 대비 훨씬 다양할 수 있다는 점에서 문제가 더 어려워진다[5]. 학습에 사용할 수 있는 공개 데이터셋 또한 다른 패션 문제를 해결하기 위한 데이터셋 대비 적다.

한편, 패션 스타일 관련 이미지 데이터 수집 시 유행하는 스타일과 그렇지 않은 스타일의 이미지 수량 차이가 발생하는 것도 모델 학습을 어렵게 하는 요인이다. 흔히 클래스 불균형 문제(Class imbalance problem)라고 명명되는 이것은 기계학습을 어렵게 하는 요인으로, 학습 후 평가 시 수량이 부족한 클래스에 속한 데이터는 다른 클래스 대비 성능이 떨어지는 문제가 발생한다. 이는 패션 관련 데이터셋에서 전반적으로 발견되는 현상으로, 데이터 수집의 규모가 커질수록 클래스 불균형 문제는 심해진다[6][7]. 위와 같은 문제들로 인하여 패션 스타일 분류는 그 중요도에 비해 연구가 부족하고, 어려운 상황이다.

이에 본 논문에서는 클래스 불균형 상황에서 패션 스타일 분류 성능 향상을 위한 모델을 제안한다. 제안 모델은 VAE(Variational AutoEncoder)[11]를 활용하여 구현되었다. VAE는 잠재 변수를 샘플링 한 후, 이를 출력 값으로 변환시키는 네트워크 구조를 가진다. 노이즈를 사용하여 샘플링을 하기 때문에 같은 입력으로부터 다양한 잠재 변수를 생성하는 것이 가능하다. 따라서 학습 이미지 수가 부족한 클래스에 대해서는 더 많은 잠재 변수를 만들어 이를 바탕으로 분류기를 학습시키는 것이 가능하다.

본 논문의 기여도를 나열하면 다음과 같다. 첫째, VAE를 활용하여 클래스 불균형 상황에서 패션 스타일을 학습하는 분류 모델을 제안하였다. 둘째, 공개 데이터셋을 통해 제안 모델이 성능을 향상시켰음을 보였다. 특히 학습 이미지 수가 적었던 클래스에서 정확도가 크게 개선되었음을 확인할 수 있었다.

본 논문의 구성을 다음과 같다. 2장은 관련 연구로 패션 스타일 관련 인공지능 연구 및 데이터셋에 대해 기술한다. 또한 데이터 수가 많은 클래스와 데이터 수가 적은 클래스를 동시에 사용하여 모델을 학습시키는 GFSL(Generalized Few-Shot Learning)[12]과 본 연구에서 활용한 방법론인 VAE 관련 연구 동향에 대해 논한다. 3장에서는 본 논문에서 제안하는 모델에 대해 자세하게 기술한다. 4장에서는 실험을 위한 데이터셋 설정 및 실험 결과에 대해 설명한다. 마지막으로 5장에서는 결론으로, 본 연구가 가지는 의의 및 한계점에 대해 논하고 향후 연구에 대해 기술한다.

Ⅱ. 관련 연구

2.1 패션 스타일 데이터셋 및 인공지능 연구

Facebook, Instagram 같은 SNS(Social Network Service)의 발전으로 사람들은 각자의 패션 스타일을 타인에게 적극적으로 알릴 수 있게 되었고, 이에 대한 관심도 이전보다 크게 증가하였다. 그럼에도 불구하고 패션 스타일 분류 문제에 대한 인공지능에 연구는 서론에서 언급한 문제들로 인하여 제한적이었다.

Takagi[8]는 14개의 클래스로 이루어진 패션 스타일 데이터셋을 공개하면서 CNN 기반 backbone 네트워크를 활용한 스타일 분류 성능을 제시하였다. 이때 [13]을 활용한 시각화 자료를 함께 제시함으로써 모델의 설명력을 표현하였다.

Miyamoto[9]는 웹사이트에서 수집한 이미지 및 클래스를 바탕으로 WEARSTYLE이라는 데이터셋을 제안하였다. 또한 이미지에서 사람을 제외한 배경을 삭제함으로써 스타일 분류 성능을 향상시킬 수 있음을 보였다.

Kiapour[10]는 5개의 클래스로 구성된 데이터셋 Hipster Wars를 제안하였다. 각 이미지 별로 얼마나 해당 스타일을 잘 표현했는지 순위가 매겨져 있는 것이 다른 데이터셋과의 차이점이라 할 수 있으며 머신러닝 기법을 사용하여 각 클래스별 성능을 평가하였다.

위와 같은 패션 스타일 관련 인공지능 연구들은 데이터셋을 제시하면서 기준 모델의 성능을 제시하는 수준에 그쳤기 때문에 모델이 고도화되지 못했다. 또한 대규모의 패션 스타일 데이터를 수집하면서 발생할 수 있는 클래스 불균형 문제를 고려하지 않고, 소규모의 클래스 균형 데이터셋만으로 모델 평가를 진행했기 때문에 실제 산업으로의 응용에 한계가 있다.

2.2 GFSL

FSL(Few-Shot Learning)이 적은 수의 데이터, 레이블 쌍만 가지고 모델을 학습시키는 것을 의미한다면, GFSL은 FSL를 확장한 개념으로 데이터가 많이 존재하는 클래스와 적은 클래스를 모두 활용하는 모델 학습 방법이다[12]. GFSL 상황에서의 성능을 측정하기 위하여 [12]의 저자들은 평가 척도로 데이터가 많은 클래스의 정확도와 적은 클래스의 정확도 사이의 조화평균을 사용하였다. 그들은 이미지와 레이블 모달리티를 모델 입력 값으로 함께 사용하여 VAE 기반 제안 모델을 학습시켰고, 제안 모델이 이미지 분류 성능을 향상시켰음을 보였다.

Huang[14]은 [12]의 모델을 기반으로 클래스 내부 잠재 변수들의 분산을 줄일 수 있는 모델을 제안하였다. 저자들은 [12]의 손실함수에 K-means 군집화 알고리즘을 활용한 식을 추가함으로써 클래스 내부 잠재 변수들의 분산을 줄였고, 이를 바탕으로 기존 모델 대비 질적으로 뛰어난 잠재 공간을 학습시킬 수 있었다고 주장하였다. 또한 향상된 성능을 통해 이를 입증하였다.

Ye[15]는 데이터가 많은 클래스로 학습시킨 분류기와 데이터가 적은 클래스로 학습시킨 분류기를 합성하여 최종적으로 GFSL 분류를 수행하는 framework를 제안하였으며, 비교 모델 대비 우수한 성능을 기록하였다.

Xian[16]은 이미지 도메인에서 수행되던 GFSL 문제를 비디오 분류 문제로 확장하였다. 시간 정보를 학습하기 위하여 spatiotemporal CNN와 3D CNN을 발전시켰으며, 제안 모델이 정확도와 검색 성능 측면에서 모두 기준 모델 대비 우수한 성능을 보임을 입증하였다.

2.3 VAE

VAE[11]는 encoder와 decoder 구조를 가진 신경망이다. Encoder를 통해서 입력 값으로부터 잠재 변수 확률 분포의 모수를 추정하여 추정된 모수로부터 잠재 변수를 샘플링하고 이를 decoder에 투입하여 입력 값을 복원하는 구조를 가지고 있다. 이 때, 확률 분포로는 일반적으로 정규분포를 가정하여 손실함수에 포함되어 있는 쿨백-라이블러 발산(Kullback-Leibler divergence)를 통해 잠재 변수 확률 분포를 정규 분포에 가깝게 근사하게 된다. 이를 종합한 손실함수는 다음과 같다.

$$\mathcal{L} = E_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x) \| p_\theta(z)) \quad (1)$$

입력을 x, 잠재 변수를 z라고 할 때 식 (1) 우변의 첫 번째 항은 입력 값의 복원 손실(Reconstruction error)을 의미한다. 파라미터 ϕ를 갖는 encoder로부터 z의 확률 분포 $q_\phi(z|x)$를 만들고, 이로부터 z를 샘플링하여 파라미터 θ를 갖는 decoder를 통해 x를 복원하여 손실을 계산하게 된다. 이 때, 역전파를 통한 학습이 가능하도록 샘플링 시 reparameterization trick을 사용하게 된다. 식 (1) 우변의 두 번째 항은 쿨백-라이블러 발산을 의미하는 부분으로, 잠재 변수의 확률 분포를 정규 분포에 가깝게 근사하는 식이다.

[11]의 저자들은 MNIST와 Frey Face 데이터셋으로 VAE를 학습시켰으며, 매니폴드의 시각화를 통해 제안한 VAE 모델이 입력 값의 정보를 잘 학습하였음을 보였다. 이후 VAE를 활용한 많은 연구들이 등장하였는데, 그 중 대표적인 것이 conditional VAE[17]와 adversarial autoencoder[18]이다. [17]의 저자들은 지도학습이 가능하도록 기존 VAE의 구조의 encoder와 decoder에 정답 레이블을 추가하였다. [18]에서의 VAE는 GAN의 generator로서 역할을 수행한다. 즉, VAE에서 샘플링된 잠재 변수가 거짓 잠재 변수가 되어 실제 데이터의 분포로부터 얻어진 참 잠재 변수와 함께 discriminator의 입력으로 사용되는 구조이다. 이를 통해 저자들은 VAE의 사전확률 분포가 표준정규분포 같은 간단한 확률 분포여야하는 단점을 극복하고자 했다.

한편, [12]의 저자들은 사용 모달리티 별로 VAE를 학습시켰다. 이 때, 모달리티 별 잠재 변수 확률 분포들이 서로 가까워지도록 하는 손실 함수를 추가하여 각각의 VAE에서 만들어지는 잠재 변수에 다른 모달리티의 정보가 포함되도록 유도하였다. 또한 VAE가 잠재 변수를 샘플링할 수 있다는 장점을 활용하여 GFSL에서 데이터 수가 적은 클래스의 잠재 변수를 데이터 수가 많은 클래스의 잠재 변수보다 많이 샘플링 하여 분류기를 학습시켰을 때 성능이 향상됨을 보였다.

III. VAE 잠재 변수를 활용한 패션 스타일 분류 모델

그림 1에서 설명되어 있는 것처럼, 제안 모델은 3단계에 걸친 모델 학습 구조를 가진다. 3단계 학습 구조는 그림 1의 (a)처럼 backbone 네트워크를 학습하는 사전 학습 단계와 그림 1의 (b)와 같이 backbone 네트워크에서 이미지의 특징을 추출한 벡터를 바탕으로 VAE를 학습하는 단계, 마지막으로 그림 1의 (c)처럼 VAE의 잠재 벡터와 backbone의 이미지 특징 벡터를 합친 벡터로 분류 레이어를 학습하는 단계로 이루어진다. 테스트 상황에서는 그림 1의 (c)에서 제시된 모델을 거쳐 생성된 출력 값을 바탕으로 스타일을 분류하게 된다. 이어 나오는 3.1, 3.2, 3.3을 통해 각각의 단계에 대해서 자세히 설명한다.

3.1 Backbone 네트워크 학습

우선, backbone 네트워크는 저수준에서 이미지의 보편적인 특징을 추출할 수 있도록 ImageNet[19]으로 사전 학습된다. 그 후, 패션 스타일 데이터셋으로 미세조정(Fine-tuning)되어, 고수준에서 패션 스타일 분류 문제에 특화된 정보를 추출할 수 있도록 한다.

이미지 분류를 위해 흔히 사용되는 여러 backbone 네트워크 중, [8]에서 최고 성능을 기록한 ResNet-50[20] 구조를 채택하여 학습을 진행한다. 다섯 번째 convolution block을 거쳐서 나온 7×7×2048의 feature를 global average pooling을 사용하여 2048차원의 벡터 x를 변환하고, 이를 fully-connected 레이어에 통과시켜 클래스 개수만큼 출력 값을 만든다. 그렇게 만들어진 출력 값과 스타일 클래스 정답으로부터 cross entropy 함수를 통해 손실 값을 계산하고, 이를 역전파하여 backbone 네트워크를 학습시킨다.

3.2 VAE 학습

VAE의 입력 값으로는 3.1에서 학습된 backbone의 이미지 feature x가 사용된다. 손실 함수로는 식 (1)의 쿨백-라이블러 발산 항에 β 가중치를 곱한 함수를 사용하여 잠재 변수가 이미지 복원 정보를 최대한 보존할 수 있도록 한다[21]. 이를 수식으로 표현하면 다음과 같다.

$$\mathcal{L} = E_{q\phi(z|x)}[\log p_\theta(x|z)] - \beta D_{KL}(q_\phi(z|x) \| \ p_\theta(z)) \quad (2)$$

본 연구에서는 z로 128차원의 벡터를 사용하며, backbone 네트워크 파라미터들은 손실에 의해서 갱신되지 않도록 고정시켜 학습의 안정성을 높인다.

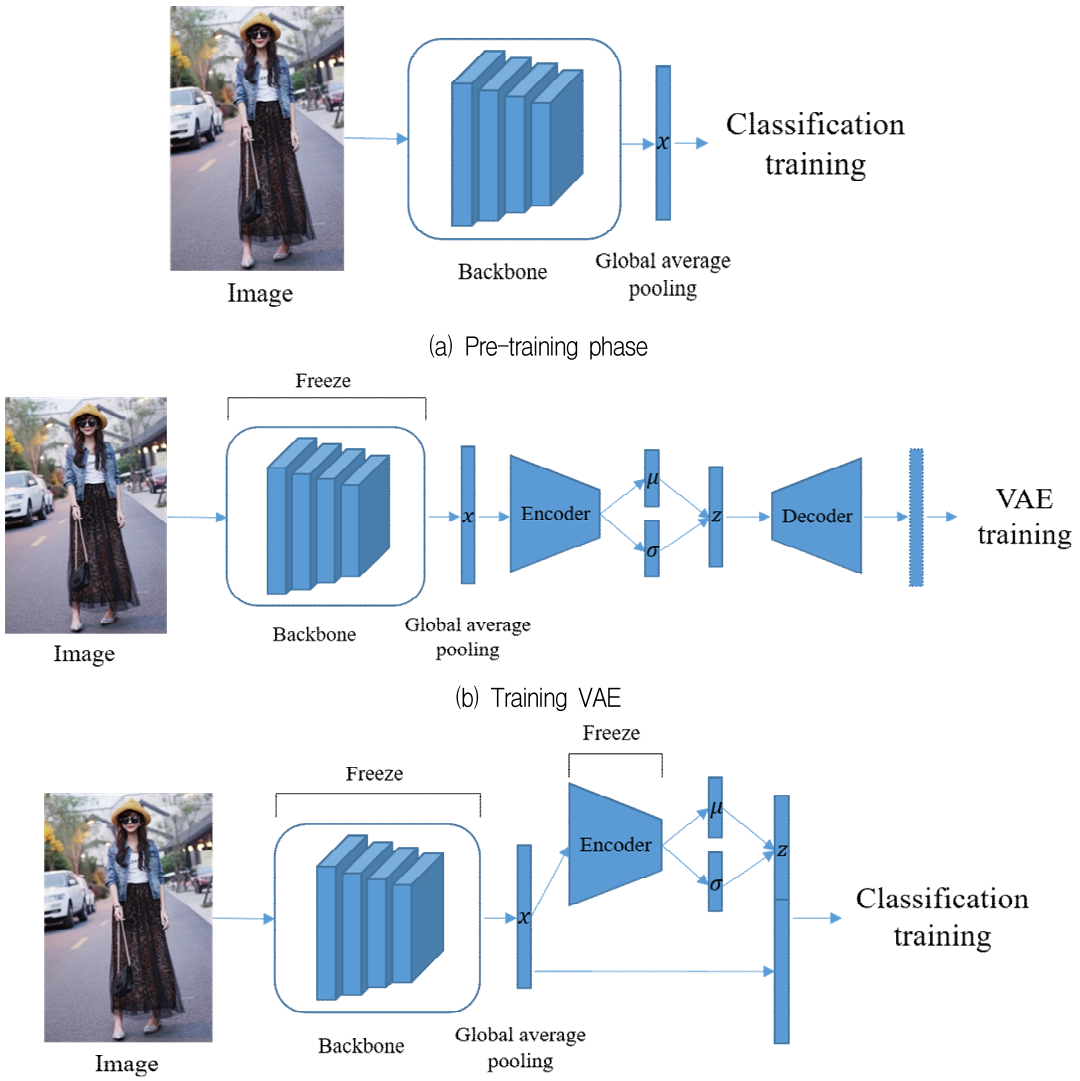

그림 1. VAE 잠재 변수를 활용한 패션 스타일 분류 모델 구조
Fig. 1. Model structure of fashion style classification using VAE latent variables

3.3 최종 분류기 학습

최종 분류기 학습을 위해서 3.1에서 학습된 backbone과 3.2에서 학습된 VAE의 encoder를 사용한다. 손실을 계산하기 위해 VAE의 encoder로부터 샘플링되는 z를 backbone의 이미지 feature x와 결합하여 fully-connected 레이어에 통과시켜 각 클래스에 대한 logit을 만들어낸다. 3.1과 같이 cross entropy를 통해 logit과 클래스 정답 사이의 손실을 계산하였으며 이를 역전파하여 네트워크를 학습한다. 이 때, backbone과 VAE의 encoder의 파라미터들은 손실에 의해서 갱신되지 않도록 고정시킨다.

IV. 실 험

4.1 FashionStyle14 데이터셋

제안 모델의 학습과 평가를 위해서 [8]에서 제안된 공개 데이터셋, FashionStyle14을 사용하였다. FashionStyle14는 총 14개의 클래스, 13,126장의 이미지로 구성되어 있다. [8]에서 제공하는 학습, 모델 평가, 테스트 데이터셋 구성이 있으나 파일이 손상되거나 중복된 경우가 많아 사용할 수 있는 파일들을 대상으로 데이터셋을 재구성하였다. 각 클래스별 이미지 수는 표 1에 나타나 있다.

표 1. FashionStyle14 데이터셋 통계
Table 1. Statistics of FashionStyle14 dataset

Class \ Phase	Training	Validation	Test	Total
conservative	535	45	312	892
dressy	501	43	292	836
ethnic	508	43	296	847
fairy	565	48	330	943
feminine	474	41	276	791
gal	558	47	325	930
girlish	656	56	382	1,094
kireime-casual	623	53	363	1,039
lolita	625	53	364	1,042
mode	628	54	366	1,048
natural	505	43	294	842
retro	499	42	291	832
rock	483	41	282	806
street	610	52	355	1,017
Total	7,770	661	4,528	12,959

GFSL 문제 상황에서 실험을 진행하기 위하여 [12] 및 [22]의 방법과 유사하게 학습용 데이터셋과 모델 평가용 데이터셋의 일부 클래스를 샘플링하였다. 구체적으로, 무작위 5개의 클래스를 선정, 그 클래스는 작은 수(n_f)의 데이터로만 구성되도록 샘플링하였다. 결과적으로, 많은 수의 이미지 데이터로 구성된 9개의 majority 클래스와 작은 수의 이미지 데이터로 구성된 5개의 minority 클래스를 합해서 학습용 데이터셋과 모델 평가용 데이터셋을 재구성하였다. 이러한 과정은 그림 2에 도식화되어 있다.

4.2 모델 학습

제안 모델의 backbone으로는 [8]에서 최고 성능을 기록한 ResNet-50 모델이 사용되었다. ResNet-50은 ImageNet으로 사전 학습되었으며, 재구성된 GFSL 문제용 FashionStyle14 데이터셋으로 미세 조정(Fine-tuning) 되었다. ImageNet으로 사전 학습된 모델은 Pytorch의 torchvision library에서 획득하였으며, 미세 조정은 50 epoch 동안 확률적 경사하강법을 통한 학습으로 진행되었다. VAE 학습은 Adam[23] 최적화 알고리즘을 사용하여 100 epoch 동안 진행되었으며, 식 (2)의 β는 90 epoch까지 매 epoch 5.6×10^{-6}의 비율로 증가하도록 스케줄링 되었다. 마지막으로 최종 분류기 학습은 10 epoch 동안 확률적 경사하강법을 통한 학습으로 진행되었다.

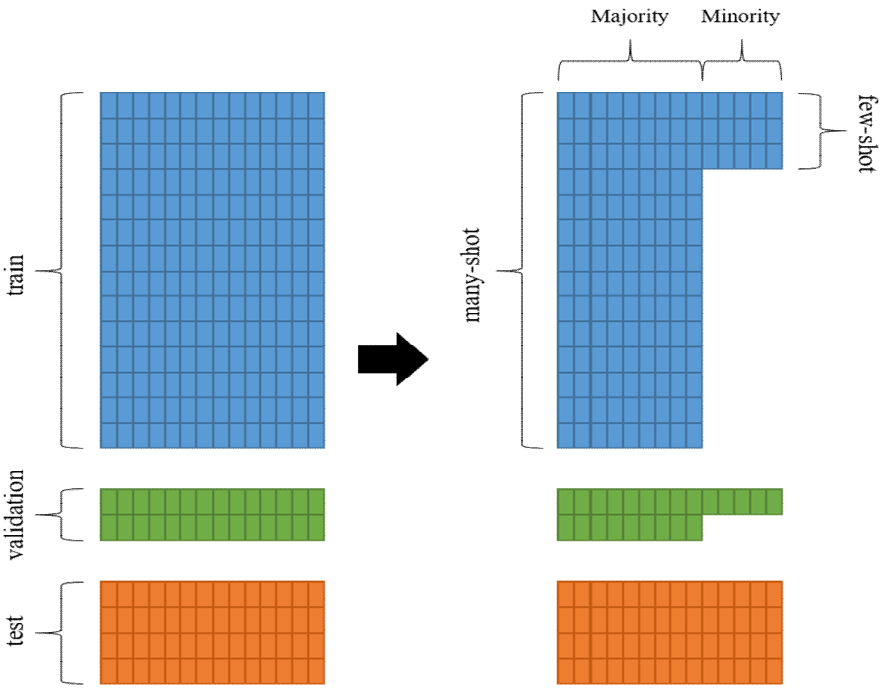

그림 2. GFSL 문제를 위한 데이터셋 구성
Fig. 2. Dataset configuration for GFSL

4.3 성능 평가

n_f의 차이에 따른 성능 변화 추이를 확인하기 위하여, 네 개의 다른 n_f (5, 10, 30, 50)를 사용하여 학습 및 평가 데이터셋을 샘플링하였다. 비교 모델로는 backbone을 미세조정한 모델과 오버샘플링을 통하여 backbone을 미세조정한 모델을 사용하였다. 오버샘플링은 minority 클래스에 속하는 데이터를 학습에 더 많이 투입하는 방법으로, 클래스 불균형 상황을 극복하기 위해 널리 사용되는 방법이다. 제안 모델의 분류기 학습 또한 VAE를 활용한 minority 클래스 오버샘플링을 통해 이루어졌다. VAE를 통한 샘플링은 확률 분포에 의한 샘플링이기 때문에 매 번 다른 값이 샘플링된다는 점에서 위의 비교 모델 오버샘플링과는 다른 특징이 있다.

그림 3은 $n_f = 10$인 데이터셋과 $n_f = 50$인 데이터셋으로 학습한 VAE의 손실 값 그래프이다. 학습이 진행됨에 따라 학습 및 모델 평가용 데이터셋에서 손실 값이 지속적으로 낮아지는 것을 확인할 수 있다. 그림 4는 $n_f = 10$인 데이터셋과 $n_f = 50$인 데이터셋에서 학습된 최종 분류기의 cross entropy 손실 값 그래프로, 학습이 진행되면서 손실 값이 하강하여 수렴하고 있는 것을 알 수 있다.

모델 성능 평가를 위한 지표로는 [12]에서 사용한 지표인 majority 클래스에 속하는 데이터들의 정확도와 minority 클래스에 속하는 데이터들의 정확도 사이의 조화 평균(acc_h)이 채택되었다. 산술 평균과는 달리, 두 정확도 사이의 곱셈을 통해 평균이 계산되기 때문에 값이 작은 minority 클래스의 정확도 비중이 커진다는 장점이 있다. 전체 데이터셋에 대한 정확도(acc) 또한 함께 측정하여 전반적인 분류기 성능도 같이 검증하였다. 테스트 데이터셋으로 성능을 측정할 모델은 매 epoch 학습 후 모델 평가 데이터셋으로부터 측정된 정확도(acc) 지표를 통해 선택되었다.

성능 측정 결과는 표 2에 기록되어있다. 전체 테스트 데이터셋의 정확도를 측정한 지표에서, 제안 모델은 비교 모델 대비 우수한 성능을 나타내었다. 이는 minority 클래스에 대한 정확도가 비교 모델보다 뛰어나기 때문이다.

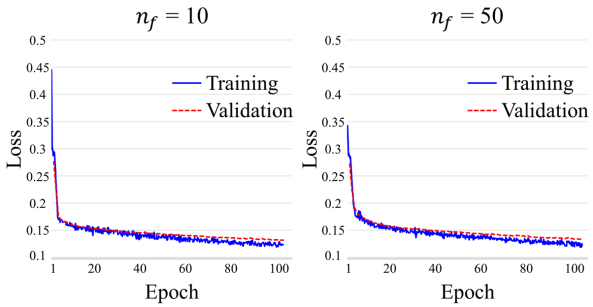

그림 3. VAE 학습 단계 별 손실 값 그래프
Fig. 3. Learning curves of the VAE loss function

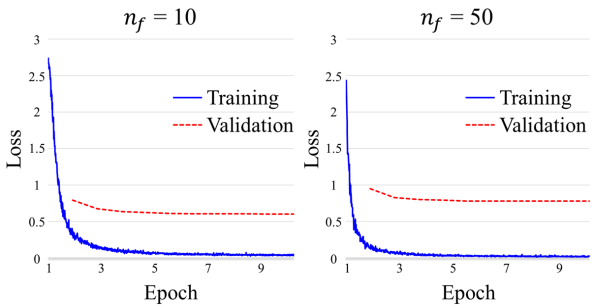

그림 4. 최종 분류기 학습 단계 별 손실 값 그래프
Fig. 4. Learning curves of the classifier's loss function

표 2. 제안 모델과 비교 모델의 성능 평가 (%)
Table 2. Performance evaluation of the models (%)

	$n_f = 5$		$n_f = 10$		$n_f = 30$		$n_f = 50$	
	acc_h	acc	acc_h	acc	acc_h	acc	acc_h	acc
ResNet-50	0.00	53.75	3.44	52.76	31.66	60.09	58.54	65.66
ResNet-50 (Oversampling×30)	4.74	54.42	12.83	54.53	29.07	59.28	58.44	65.50
ResNet-50 (Oversampling×50)	4.01	54.28	11.89	54.00	28.61	59.28	59.78	66.43
Proposed model (VAE sampling×30)	3.64	54.06	31.34	58.52	51.42	65.33	66.01	68.20
Proposed model (VAE sampling×50)	6.18	54.59	37.24	60.00	51.28	65.35	66.39	68.38

VAE 학습을 통해 얻어진 잠재 변수의 확률 분포가 이미지의 정보를 잘 내포하고 있어서, 이로부터 샘플링된 잠재 변수가 minority 클래스에 대해서 데이터 증강(Data augmentation) 효과를 가져온 것으로 파악된다. 이는 오버샘플링 방식과의 비교에서 잘 나타나는데, 적은 수의 이미지가 매번 같은 feature를 만들어 minority 클래스에 대해서 과적합되기 쉬운 오버샘플링 방식과 달리 VAE로부터 얻어진 잠재 변수는 같은 이미지라도 매번 다른 feature를 만들기 때문에 모델이 과적합되지 않는다.

한편, VAE 잠재 변수로 인한 성능 향상 효과는 $n_f = 10$ 또는 $n_f = 30$일 때 가장 컸다. $n_f = 5$일 때는 minority 클래스의 데이터 수가 너무 적어서 VAE의 잠재 변수가 유의미한 정보를 포함하지 못하였고, $n_f = 50$ 일 때는 비교 모델도 minority 클래스에 대해서 학습이 가능할 만큼의 데이터 수량이 확보되어 제안 모델과의 편차가 줄어든 것으로 해석되어진다.

V. 결론 및 향후 과제

본 연구에서는 VAE를 활용하여 클래스 불균형 상황에서 패션 스타일 분류 성능을 향상시키는 모델 학습법에 대해 다루었다. 상황 설정을 위하여 GFSL 문제를 도입하였으며, 학습을 위한 데이터셋을 생성하기 위해 패션 스타일 관련 공개 데이터셋인 FashionStyle14을 재구성하였다. 실험 결과, 제안 모델이 비교 모델 대비 정확도와 전반적 분류기 성능 측면에서 더 나은 결과를 보이는 것을 확인할 수 있었다. 특히, 제안 모델은 오버샘플링 방식과의 비교에서 minority 클래스 정확도의 비중을 높인 acc_h 지표를 네 가지 데이터셋($n_f = 5, 10, 30, 50$)에 대해 각각 평균적으로 0.54%, 21.93%, 22.41%, 7.09% 개선시켰다.

클래스 불균형 상황에서 성능을 더욱 향상시키기 위하여 다른 기존 이미지 처리 관련 연구를 패션 스타일 분류 문제로 확장시켜 보는 것은 본 연구의 향후과제라고 할 수 있다. 특히, 제안 모델은 이미지만 사용하여 VAE를 학습시켰으나 [12]에서처럼 다양한 모달리티를 활용하여 VAE 잠재 변수의 확률 분포에 더 많은 정보가 포함되도록 한다면, 성능은 더욱 개선될 것이다.

FashionStyle14가 아닌 다른 데이터셋으로 제안 모델을 검증해보는 것도 의미가 있을 것이다. FashionStyle14는 일본에서 제작되었기 때문에 국내 및 다른 국가의 패션 스타일과는 차이가 있는 부분이 있다. 따라서 패션 스타일 관련 이미지를 새로 수집하여 모델을 학습시키고, 성능을 측정해보는 것 또한 국내외 패션 산업에서의 활용 가능성 측면에서 의미가 있을 것이다.

References

[1] Y. J. Nam and H. H. Jo, "Prediction of Weekly Load using Stacked Bidirectional LSTM and Stacked Unidirectional LSTM", Journal of Korean Institute of Information Technology, Vol. 18, No. 9, pp. 9-17, Sep. 2020.

[2] S. W. Jung, H. J. Kwon, Y. C. Kim, S. H. Ahn, and S. H. Lee, "Image Translation Method based on GAN for Surveillance under Dim Surround", Journal of Korean Institute of Information Technology, Vol. 18, No. 8, pp. 9-17, Aug. 2020.

[3] M. H. So, C. S. Han, and H. Y. Kim, "Defect Classification Algorithm of Fruits Using Modified MobileNet", Journal of Korean Institute of Information Technology, Vol. 18, No. 7, pp. 81-89, Jul. 2020.

[4] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network", Advances in neural information processing systems (NeurIPS), Denver, CO, USA, pp. 396-404, Nov. 1990.

[5] G. L. Sun, X. Wu, H. H. Chen, and Q. Peng, "Clothing Style Recognition using Fashion Attribute Detection", the 8th International Conference on Mobile Multimedia Communications, Chengdu, China, pp. 145-148, May 2015.

[6] Q. Chen, J. Huang, R. Feris, L. M. Brown, J. Dong, and S. Yan, "Deep Domain Adaptation for Describing People based on Fine-grained Clothing Attributes", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 5315-5324, Jun. 2015.

[7] Z. Liu, P. Luo, S. Qiu, L. X. Wang, and X. Tang, "Deepfashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 1096-1104, Jun. 2016.

[8] M. Takagi, E. Simo-Serra, S. Iizuka, and H. Ishikawa, "What Makes a Style: Experimental Analysis of Fashion Prediction", The IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, pp. 2247-2253, Oct. 2017.

[9] R. Miyamoto, T. Nakajima, and T. Oki, "Accurate Fashion Style Estimation with a Novel Training Set and Removal of Unnecessary Pixels", The IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Hokkaido, Japan, pp. 1-5, May. 2019.

[10] M. H. Kiapour, K. Yamaguchi, and A. C. Berg, and T. L. Berg, "Hipster Wars: Discovering Elements of Fashion Styles", The 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, pp. 472-488, Sep. 2014.

[11] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes", arXiv:1312.6114, Dec. 2013.

[12] E. Schonfeld, S. Ebrahimi, S. Sinha, T. Darrell, and Z. Akata, "Generalized Zero-and Few-shot Learning via Aligned Variational Autoencoders", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 8247-8255, Jun. 2019.

[13] R. C. Fong and A. Vedaldi, "Interpretable Explanations of Black Boxes by Meaningful Perturbation", The IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 3429-3437, Oct. 2017.

[14] Y. Huang, Z. Deng, and T. Wu, "Learning Discriminative Latent Features for Generalized Zero-and Few-Shot Learning", The IEEE International Conference on Multimedia and Expo (ICME), London, UK, pp. 1-6, Jul. 2020.

[15] H. J. Ye, H. Hu, D. C. Zhan, and F. Sha, "Learning Adaptive Classifiers Synthesis for Generalized Few-Shot Learning", arXiv:1906.02944, Jun. 2019.

[16] Y. Xian, B. Korbar, M. Douze, B. Schiele, Z. Akata, and L. Torresani, "Generalized Many-Way Few-Shot Video Classification", arXiv:2007.04755, Jul. 2020.

[17] K. Sohn, H. Lee, and X. Yan, "Learning Structured Output Representation using Deep Conditional Generative Models", Advances in neural information processing systems (NeurIPS), Montreal, QC, Canada, pp. 3483-3491, Dec. 2015.

[18] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial Autoencoders", arXiv:1511.05644, Nov. 2015.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks", Advances in neural information processing systems (NeurIPS), Lake Tahoe, NV, USA, pp. 1097-1105, Dec. 2012.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778, Jun. 2016.

[21] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework", International Conference on Learning Representations (ICLR), Toulon, France,

pp. 1-22, Apr. 2017.

[22] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly", The IEEE transactions on pattern analysis and machine intelligence (TPAMI), Vol. 41, No. 9, pp. 2251-2265, Sep. 2019.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", International Conference on Learning Representations (ICLR), San Diego, CA, USA, pp. 1-15, May 2015.

저자소개

박 종 혁 (Jonghyuk Park)

2015년 2월 : 서울대학교 산업공학과(공학사)
2015년 1월 ~ 2016년 2월: ㈜삼성전자 메모리 사업부
2016년 3월 ~ 현재 : 서울대학교 산업공학과(석박사통합과정) 재학 중

관심분야 : 컴퓨터 비전, 딥러닝, 머신 러닝 응용