

일반논문 (Regular Paper)

방송공학회논문지 제29권 제1호, 2024년 1월 (JBE Vol.29, No.1, January 2024)

<https://doi.org/10.5909/JBE.2024.29.1.1>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

객체 비율을 고려한 패션 이미지 분류 시스템

최은지^{a)}, 윤경로^{a)†}

Fashion Image Classification System Considering Object Aspect Ratio

Eunji Choi^{a)} and Kyoungro Yoon^{a)†}

요약

최근 하드웨어 기술의 발달로 딥러닝 기술은 연산량 문제와 연산 속도를 개선하며 다양한 분야에 적용되고 있다. 이러한 변화에 따라 각 분야에서 객체 인식의 정확도 성능을 높이려는 시도가 활발히 연구되고 있다. 본 논문에서는 객체의 비율이 중요하게 평가되는 패션 이미지 인식 시스템을 위해 제로 패딩과 보더 외삽법을 이용한 객체 비율 보존 방법에 기반한 패션 이미지 분류를 제안한다. 제안하는 방법은 Confusion Matrix의 위양성률과 위음성률을 감소시켜 오분류 문제를 해결하고, 제안 방법의 효과를 통해 비율이 중요한 패션 이미지 인식에서 성능 향상을 보여준다.

Abstract

With the recent development of hardware technology, deep learning technology has improved computational volume problems and computational speed, and is being applied to various fields. According to these changes, attempts to increase the accuracy performance of object recognition in each field are being actively studied. In this paper, Object ratio preservation method is proposed using zero padding and border extrapolation for a fashion image recognition system in which ratio of objects is evaluated important. It solves the problem of misclassification by reducing false positives rate and false negatives rate of the Confusion matrix, and shows performance improvement in fashion image recognition where ratio is important through effect of the proposed method.

Keyword : Deep learning, CNN (Convolutional Neural Network), Fashion image classification

a) 건국대학교 스마트ICT융합공학과(Smart ICT Convergence Engineering Dept. Konkuk Univ.)

† Corresponding Author : 윤경로(Kyoungro Yoon)

E-mail: yoonk@konkuk.ac.kr

Tel: +82-2-450-4129

ORCID: <https://orcid.org/0000-0002-1153-4038>

· Manuscript September 1, 2023; Revised January 25, 2024; Accepted January 26, 2024.

1. 서론

최근 딥러닝 기술은 ‘AI의 겨울’을 지나 하드웨어의 발달로 인하여, 연산속도가 개선되면서 활발히 연구되고 있다. 이에 객체인식과 관련한 자율 주행, 의료, 국방, 의류 등 다양한 분야에서 신경망을 이용한 자동화 시스템을 구축하기 위한 시도가 진행되고 있다. 딥러닝 기술을 적용한 패션 이미지 인식 시스템은 이미지 검색을 기반으로 한 추천시스템과 상품 분류 등 다양한 플랫폼에 적용되고 있으며, 패션 이미지 객체 인식 기술에서 있어 객체의 스타일과, 재질, 사이즈, 패턴 등의 요소는 패션 이미지 분류에 있어 중요한 요소로 평가되고 있다. 이에, 다양한 특성을 가진 의류의 분류는 효과적인 이미지 특성 추출방식이 요구된다. 본 논문에서는 CNN(Convolutional Neural Network) 학습에서, 다양한 해상도의 입력 이미지가 정해진 사이즈의 입력 이미지로 맞춰지는 과정에서 발생하는 픽셀의 왜곡현상을 해결하기 위해 객체 비율 보존 기법을 제안한다. 본 논문에서 1장과 2장에서는 서론과 배경지식 및 관련 연구를, 3장에서는 객체 비율 보존 기법을 적용한 시스템 구성도를 설명하고, 4장에서 실험 설계 및 구성과, 5장 결론 및 향후 과제를 통해, 본 연구의 활용성을 서술한다.

II. 배경 지식 및 관련 연구

1. CNN(Convolutional Neural Network)

AI의 겨울의 시기인 1960~1990년에는 AI와 관련된 연구가 더디었으나 2000년대부터 점진적으로 AI와 관련된 연구가 진행되었으며 2010년을 기점으로 AI와 관련된 기술들이 급속도로 발전하기 시작한 모습을 보여준다. 객체 분류 기법이란 이미지에 존재하는 특정 객체가 무엇인지 또는 어떤 종류인지 분류하는 기술로, 입력 이미지에 커널 필터가 지나며, 컨볼루션 연산을 통해 특징 맵을 추출하고 각 레이어에서 얻은 특징 점을 통해 고차원 이미지를 인식하는 기술을 말한다^[1]. 현재 CNN의 오류율은 2010년도 28%에서 현재 2.25%로 인간이 인식하는 수준을 뛰어넘으며, CNN의 성능은 90%의 정확도에 도달하였다. CNN

기반 객체 분류 모델에는 AlexNet^[17], VGGNet^[18], ResNet, DenseNet 등의 모델이 있다.

ResNet은 2015년도 ILSVRC에서 우승을 차지한 CNN 신경망 모델로, 152층의 layer로 구성된 Skip Connection 구조의 네트워크이다^[2]. 기울기 소실(Vanishing gradient) 문제란, 네트워크가 깊어질수록 역전파과정에서 기울기가 소실되는 문제를 말한다. ResNet은 VGGNet 이후, CNN 신경망의 네트워크가 깊어짐에 따라 나타나는 기울기 소실 문제를 극복하고자 Skip Connection 구조를 도입하였다. ResNet은 Skip Connection를 통해, short cut 레이어를 만들어 컨볼루션을 수행하여 나온 가중치에 identity block을 거쳐 나온 가중치를 더해주는 방식으로 문제를 해결하였다. ResNet의 전체적인 구조는 여러 개의 residual block을 반복적으로 쌓아 구성하였으며, 네트워크의 구조를 깊게 쌓으면서도 기울기 소실 문제를 해결하여 신경망 학습의 성능을 증가시켰다.

DenseNet은 2017 CVPR 컨퍼런스에서 발표된 논문으로, ResNet과 같이 CNN의 layer가 깊어지면서 기울기 소실 문제를 해결하고 신경망 구조의 효율화를 제안된 모델이다^[3]. DenseNet은 Dense connectivity라는 개념을 통해 연속적인 모든 layer들에 대해 직접적인 연결을 수행한다. Dense connectivity는 채널단위로 특성맵을 이어붙여, 초기 정보들이 하나의 블럭 마지막 부분까지 전달될 수 있도록 만든다. DenseNet 한 블럭 내 모든 레이어를 이어붙여 직접 연결 방식을 채택하여 연산량은 한층 줄어들면서 특성 맵을 효율적으로 전달하는 방식을 구조를 제안하였다. DenseNet의 전체적인 구조는 Dense block 사이사이에 1x1 컨볼루션과 2 x 2 average pooling으로 구성된 transition layer가 된 삽입된 구조를 가져 ResNet과 비등한 성능을 보이나 파라미터 수는 ResNet이 가진 파라미터 수의 약 1/3 수준으로 줄었다.

2. 입력 이미지 영상 크기 조절 방식

고유한 객체 특성을 가지고 있는 이미지 분류를 위해, CNN 신경망의 학습에서 이미지의 특징점 뿐만 아니라, 질감과 위치 정보, 비율 정보 등이 필요하다^{[4][5]}. 또한, 학습을 위한 이미지의 비율 정보는 특정 클래스 분류에 있어 중요

한 정보로 판단된다^{[6][7]}. 이미지 영상 크기는 객체가 가진 고유한 정보를 유지하거나 이미지 스케일링을 통해 이미지 영역을 조절함으로써, 객체 인식에 있어 중요한 정보를 제공한다^[8].

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	30	27	27	27	29	0	0
0	0	28	25	26	31	28	0	0
0	0	29	22	24	30	26	0	0
0	0	24	21	25	29	25	0	0
0	0	26	25	24	28	24	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

그림 1. 패딩 연산을 이용한 입력 이미지 영상 크기 조절
 Fig. 1. Visualization of adjusting the size of the input image using zero padding

그림 1은 패딩 연산을 이용하여 입력 이미지 영상의 크기를 조절하는 과정을 시각화한 것이다. 일반적으로 패딩 연산은 이미지 크기를 유지하면서, 커널의 수행을 위한 연산 범위를 확장하는데 사용된다. 패딩 연산은 이미지의 경계를 벗어난 영역에도, 컨볼루션 연산을 가능하게 함으로써, 가장자리에 위치하는 이미지 정보의 손실을 줄이는 역할을 한다. 또한, 이것은 데이터 전처리와 데이터 증강 과정에서 입력 이미지의 주변을 가상의 픽셀로 할당하여 이미지 크기를 유지하게 기능함으로써, 절대적인 이미지 정보를 보존하고 이미지의 공간 정보를 인코딩하는 역할을 수행한다^[4]. 이것은 스트라이드(Stride)와 함께, 출력 특성 맵의 크기를 세밀하게 제어하는 데 활용되며, 이미지 가장자리의 픽셀에 패딩을 추가하여 입력이미지 변형에 대한 모델 민감성을 낮추는 역할을 한다.

패딩 연산 이외에도, 이미지의 공간 정보를 유지시키는 방법에는 테두리 외삽법(Border Extrapolation)이 있다^[9]. 테두리 외삽법이란 이미지 해상도 밖에 위치하는 이미지의

픽셀을 처리하는 기법의 한 종류이다^[10]. 일반적인 방식으로, 가상의 픽셀을 복제하거나 반영하여 임의의 값으로 처리하거나 가장 가장자리에 있는 픽셀 값으로 채워주는 방식 등이 있다. 주로, 이미지 처리에서 경계 픽셀 값이 필요한 영역을 처리하는 데 활용된다.

3. 패션 이미지 분류를 위한 객체 인식

딥러닝 기반의 자동화 시스템은 AI 연구의 발달과 함께, 다양한 산업 군에서 기술을 적용하기 위한 시도가 이어지고 있다. 기계 학습 기반 패션 이미지 인식에는 코사인 유사도 거리를 이용한 패션 이미지 검색^[11], 의류 랜덤 마크 탐지를 위한 객체 탐지^[12], 딥 러닝을 이용한 패션 의류 속성 인식^[13] 등의 연구가 진행되고 있으며, 패션 객체 인식 기술에 있어 객체의 질감, 스타일, 비율과 핏 등의 고유한 정보는 분류의 정확성을 위한 중요한 요소로 평가된다^[14]. 이에, 의류분야 이미지 분류에 있어 객체 이미지의 효과적인 특성 추출방식이 요구된다.

FashionSearchNet은 패션 객체의 길이와 스타일 등의 특성 인식을 이용한 패션 이미지 분류 모델이다^[15]. FashionSearchNet은 분류 손실 값(Classification loss)에 특정 속성이 두드러진 영역에 대한 Localization module을 적용함으로써 객체 속성 표현을 추출하고 유사도 학습을 통해 이미지를 분류한다. Localization module이란 Attribute Activation Maps(AAM)으로부터 얻은 활성화 이미지의 특성 경계를 가진 특징 맵을 추출하여 유사도 학습을 위해 각 fully connected layer에 특징 맵을 전달하는 역할을 한다. FashionSearchNet은 삼중항 손실 기반 유사도 학습 프로세스를 사용함으로써 유사한 객체 특성을 갖는 이미지를 분류한다.

SizeNet은 teacher-student 접근법을 통해 사이즈 특성을 반영하고 있는 패션 객체의 사이즈 분류가 가지고 있는 이슈^[15]에 대한 해결책을 제안한다^[16]. SizeNet은 이미지 분류 문제에 통계적인 접근법을 추가함으로써, 약한 지도학습(weakly-supervised learning)기반 패션 객체의 분류를 수행한다.

III. 시스템 구성도

1. 객체 비율 보존 기법을 적용한 시스템 설계도

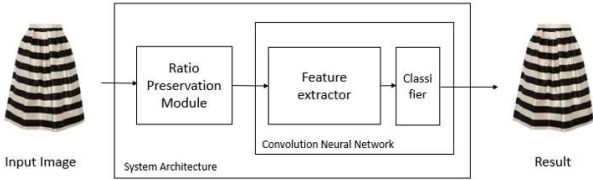


그림 2. 시스템 설계도
Fig. 2. System Architecture

그림 2는 본 논문에서 제안하는 객체 비율 보존 기법의 성능을 확인하기 위한 분류 시스템 설계도이다. 입력 이미지의 객체 비율을 보존하기 위해서 비율 보존 모듈(Ratio Preservation Module)을 이용하여 해상도 중횡비를 유지시킨 후, CNN을 통해 이미지의 특징을 추출하여 신경망 학습을 진행한다. 이후, 분류기(Classifier)를 통해 객체의 비율이 다른 유사 클래스 간의 정확도를 분류한다. 객체 비율 보존 기법을 적용한 비율 보존 모듈의 원리는 다음 절에서 설명한다.

2. 객체 비율 보존 기법

패션 데이터셋 이미지에는 다양한 해상도의 이미지가 존재하고, 객체의 특성에 따라 각기 다른 비율을 가지고 있다. 패션 영상 객체 분류에 있어 객체의 고유한 비율은 신경망 학습에서 고려해야 할 중요한 요소가 된다. 그러나, 일반적인 CNN 신경망에서는 가로, 세로 픽셀 수가 정해진 크기의 입력 이미지에 컨볼루션 연산을 하여 특징점을 추출한다^[16]. 다양한 크기의 원본 이미지를 정해진 크기의 입력 이미지로 맞춰주는 과정에서 픽셀이 생략되거나 변형되면서, 원본 이미지의 픽셀 정보의 왜곡 현상이 일어난다. 이때, 객체 중횡비의 왜곡은 일부 클래스 객체의 클래스 오분류를 야기할 가능성이 있다. 따라서, 기존 객체가 가진 이미지의 정보 손실의 방지하고, 특성 맵이 갖는 값을 제어하여 객체 중횡비의 왜곡을 방지하는 방법이 제안한다.

그림 3은 본 논문에서 제안하는 객체 비율을 보존하는 이미지 리사이즈 알고리즘이다. 본 논문에서 제안하는 객체 비율 보존 기법은 CNN 학습을 위해, 고정된 사이즈로 조정되는 과정에서 발생하는 픽셀의 왜곡 현상을 방지하고, 원본 이미지가 가지는 다양한 해상도의 중횡비 보존을 통해 객체의 비율을 보존한다. 입력 이미지가 가지는 해상도의 중횡비 보존을 위해서, 원본 이미지의 높이와 너비의 길이를 얻은 다음 높이와 너비의 크기 비교를 통해 더 높은 값을 가진 기준에 따라 이미지의 해상도를 설정한다. 이후, 최대 해상도 비로 설정한 부분과 원본 이미지의 차이(α)만큼 가상의 픽셀로 채워 이미지를 생성한다.

객체 비율 보존 기법은 목표 중횡비를 유지하는 최소 바운딩 박스의 중앙에 입력 영상을 배치하고 빈 공간을 제로 패딩 또는 테두리 외삽법을 이용하여 채운 후, 보간법을 통해 고정된 목표 크기로 조정한다. 따라서, 본 논문에서 제안하는 객체 비율 보존 기법의 경우, 주어진 이미지의 해상도에 따라 상하 또는 좌우로 패딩을 적용하게 된다.

높이가 폭보다 큰 경우는 아래 식 (1)의 방식을 따르고 그렇지 않을 경우는 아래 식 (2)의 방식을 따라서 픽셀 값을 생성하게 된다.

$$I'(x, y) = \begin{cases} pad(x, y), for 0 \leq x < \frac{\alpha}{2} \\ I(x, y), for \frac{\alpha}{2} \leq x < \max(x) + \frac{\alpha}{2} \\ pad(x, y), for \max(x) + \frac{\alpha}{2} \leq x < \max(x) + \alpha \end{cases} \quad (1)$$

$$I''(x, y) = \begin{cases} pad(x, y), for 0 \leq y < \frac{\alpha}{2} \\ I(x, y), for \frac{\alpha}{2} \leq y < \max(y) + \frac{\alpha}{2} \\ pad(x, y), for \max(y) + \frac{\alpha}{2} \leq y < \max(y) + \alpha \end{cases} \quad (2)$$

식 (1)에서는 원본 이미지의 Height 값이 Width의 값보다 높고, 가로 축의 값이 $0 \leq x < \frac{\alpha}{2}$ 또는 $\max(x) + \frac{\alpha}{2} \leq x < \max(x) + \alpha$ 구간에 존재할 때, 가상의 픽셀을 할당하여 채운 후, 보간법을 적용해 목표 해상도로 이미지 크기를 조정된 이미지를 생성한다.

식 (2)에서는 원본 이미지의 Width 값이 Height의 값보다 높고, 세로 축의 값이 $0 \leq y < \frac{\alpha}{2}$ 또는 $\max(y) + \frac{\alpha}{2} \leq x < \max(y) + \alpha$ 구간에 존재할 때, 가상의 픽셀을 할당하여 채

운 후, 보간법을 적용해 목표해상도로 이미지 크기를 조정 한 이미지를 생성한다.

3. 제로 패딩 기법(Zero padding)을 적용한 객체 비율 보존 기법

그림 4는 본 논문에서 제안하는 객체 비율 보존을 위해 제로 패딩 연산을 수행한 이미지의 예시를 보여준다. 제로 패딩 연산을 적용한 객체 비율 보존 기법은 이미지의 해상도 비를 유지한 채, 더 높은 비를 가진 기준에 따라 이미지 사이지를 설정한 후, 이미지 픽셀 값 외의 나머지 부분의 픽셀 값을 0으로 채워주는 방법을 취한다.

4. 테두리 외삽법(Border Extrapolation)을 적용한 객체 비율 보존 기법

그림 5는 객체 비율 보존 방법 중 테두리 외삽법을 적용한 이미지 예시를 보여준다. 테두리 외삽법을 적용한 객체 비율 보존 기법은 해상도 비를 유지한 채 남은 픽셀을 가장 근접한 가장자리의 픽셀을 복사하여, 채워 넣는 방법을 취한다. 이 방법은 다른 테두리 외삽법과는 달리 객체의 이미지 픽셀 값을 훼손시키지 않으면서 공간 정보를 확장하는 방식을 사용하여, 고유한 객체 이미지 정보를 유지한다.

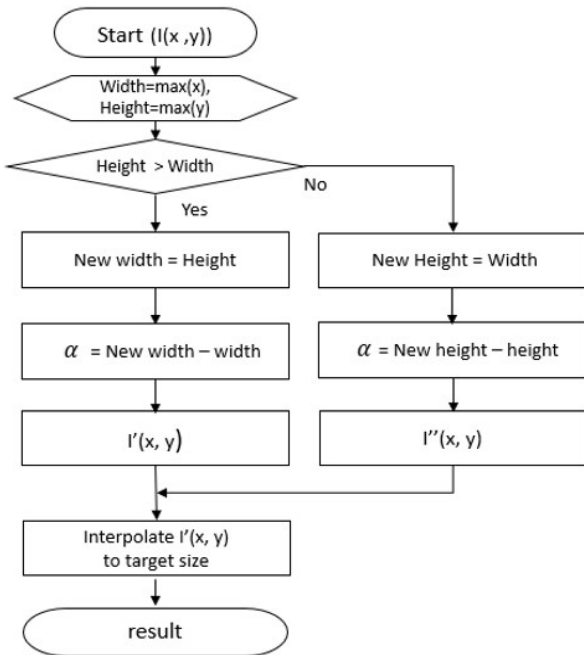


그림 3. 객체 비율 보존 기법 알고리즘 순서도
 Fig. 3. Flowchart of Object Ratio Preservation Method Algorithm

0	0	31	29	27	28	29	0	0
0	0	32	29	28	29	30	0	0
0	0	30	27	27	27	29	0	0
0	0	28	25	26	31	28	0	0
0	0	29	22	24	30	26	0	0
0	0	24	21	25	29	25	0	0
0	0	26	25	24	28	24	0	0
0	0	23	26	31	27	27	0	0
0	0	22	25	30	26	27	0	0

그림 4. 제로 패딩 기법을 시각화한 이미지
 Fig. 4. Visualized image of Zero Padding method

31	31	31	29	27	28	29	29	29
32	32	32	29	28	29	30	30	30
30	30	30	27	27	27	29	29	29
28	28	28	25	26	31	28	28	28
29	29	29	22	24	30	26	26	26
24	24	24	21	25	29	25	25	25
26	26	26	25	24	28	24	24	24
23	23	23	26	31	27	27	27	27
23	23	23	26	31	27	27	27	27

그림 5. 테두리 외삽법을 시각화한 이미지
 Fig. 5. Visualized image of Border Extrapolation method

IV. 실험의 설계 및 구성

1. 데이터셋 구성

실험에서 사용한 데이터 셋은 Deep fashion dataset의 계층적으로 구성된 클래스들을 재구성하여 사용하였으며, 비율을 포함하고 있는 클래스 간의 분류와 관계없는 클래스를 제외하였다. 또한, 테두리 외삽법을 적용한 비교 실험을 위해 가장자리에 객체가 있는 이미지는 학습에서 제외하여

사용하였다. 데이터 셋의 구성은 총 38,810장으로, Train set 23,292장, Val set 7,757장 Test set 7,761장으로 구성하였다. 또한, 데이터 수가 불균형한 클래스에 대하여 웹 크롤링을 통해 train set의 1,737장과 test set의 7,761장의 데이터를 추가하였다. 데이터셋의 레이블은 Top, Crop top, Short pants, Midi pants, Long pants, Short skirt, Midi skirt, Long skirt, Mini dress, Midi dress, Long dress, Shirt, Shirts dress의 13개 클래스로 구성하였다.



그림 6. 데이터셋 구성의 예
Fig. 6. Example of dataset configuration

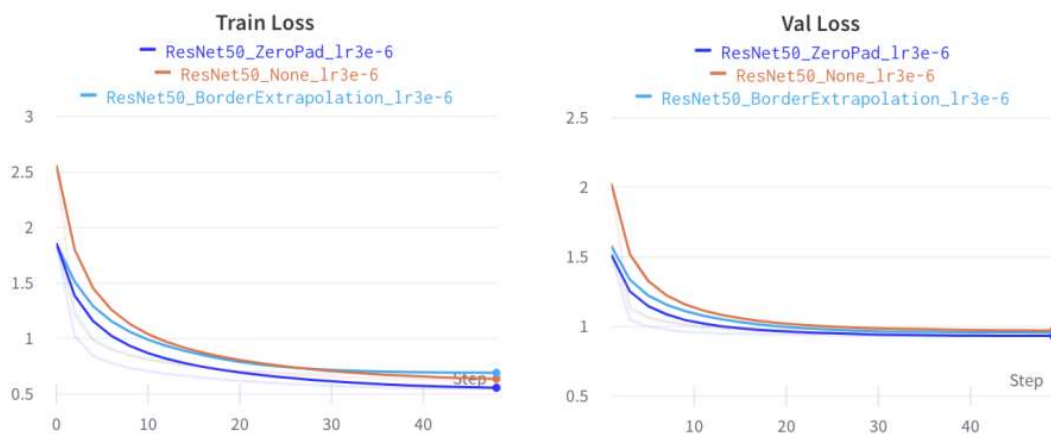


그림 7. 손실함수 그래프
Fig. 7. Loss graphs of Train and Validation

2. 실험 방법

신경망 학습은 ImageNet으로 사전 학습된 CNN 신경망에 객체가 가지는 비율을 기준으로 분류할 수 있는 Dataset을 미세 조정(Fine-tuning)하여 진행되었다. 실험에는 CNN 신경망 모델인 ResNet50과 DenseNet201을 사용하였으며, 객체 비율 보존 기법의 유효성을 입증하기 위해 동일한 파라미터를 적용한 환경에서 실험하였다.

V. 실험 결과

1. 성능 평가

그림 7은 Train set과 Validation set을 사용하여 손실을 계산할 때의 모습을 시각화한 결과다. Train set과 Validation set에서 모두 일반적인 신경망을 사용했을 때 보다, 비율 보존 모듈을 적용하여 실험하였을 때 더 적은 손실을 갖는 것을 확인할 수 있다.

표 1. 정밀도, 재현율, F1 Score 결과
 Table 1. Precision, Recall, F1 Score

Macro average	ResNet50 +None	ResNet50 +Zero padding	ResNet50 +Border extrapolation
Precision	74.31%	76.75%	76.37%
Recall	73.87%	76.40%	76.69%
F1score	74.09%	76.57%	76.53%

표 1은 ResNet50 신경망을 이용할 때 영상크기 조절 방법의 성능비교를 보여 준다. 비율 보존 모듈을 적용하지 않은 실험과 제로 패딩 모듈을 적용하여 실험한 결과 정밀도에서 약 2.44%, 재현율은 약 2.53%, F1 score에서 약 2.48% 차이를 보였다. 또한, 테두리 외삽법 모듈을 적용한 실험에서는 모듈을 적용하지 않은 실험과, 정밀도에서 약 2.06%, 재현율에서 약 2.82%, F1 score에서 약 2.44% 차이를 보였다. 위의 결과는 정밀도, 재현율, F1 score에서 전반적으로 객체 비율 보존 기법이 패션 이미지 분류 시스템 성능을 향상시키는 것을 보여준다.

표 2는 ResNet50과 DenseNet201에서의 성능을 보여준

표 2. ResNet50, DenseNet201의 Top1, Top5 정확도 비교
 Table 2. Top1, Top5 Accuracy Comparison of ResNet50 and DenseNet201

	Top1@Acc	Top5@Acc
ResNet50 +None	73.87%	91.86%
ResNet50 +Zero padding	76.40%	94.81%
ResNet50 +Border extrapolation	76.37%	94.53%
DenseNet201 +None	74.66%	92.54%
DenseNet201 +Zero padding	77.23%	95.63%
DenseNet201 +Border extrapolation	76.94%	95.26%

다. ResNet50 신경망에 객체 비율 보존 기법을 적용하지 않았을 때와 두 방법으로 객체 비율 보존 기법을 적용했을 때, Top 1 accuracy는 각각 73.87%와 76.40%, 76.37%로, 약 2.53%와 2.5%로 차이를 보였으며, Top5 accuracy는 각각 91.86%와 94.81%, 94.53%이다. 또한, DenseNet201의 경우, Top 1 accuracy는 각각 74.66%와 77.23%, 76.94%로, 약 2.57%, 2.28% 차이를 보였으며, Top 5 accuracy는 각각 92.54%와 95.63%, 95.26%이다. 위 실험에서는 각기 다른 비율을 가지고 있는 객체의 데이터 셋을 이용한 학습하여 비교한 결과, 객체 비율 보존 기법을 적용한 ResNet과 DenseNet을 활용한 실험에서 패션 이미지 분류 시스템의 성능이 향상되는 것을 알 수 있다.

표 3. 기존 신경망과 객체 비율 보존 기법을 적용한 신경망의 성능 비교
 Table 3. Network performance comparison of various resizing methods

	ResNet50 + None	ResNet50 +Zero padding	ResNet50 +Border extrapolation
TPR	74.31%	76.75%	76.37%
FPR	2.18%	1.97%	1.96%
TNR	97.82%	98.03%	98.04%
FNR	25.69%	23.25%	23.63%

표 3은 일반적인 신경망과 제안한 방법을 적용했을 때와 zero padding, border extrapolation을 이용하여 원본의 해상

도를 보존하였을 때의 confusion matrix 분석 결과^[19]이다. 세 실험에서 실제 값과 예측 값이 일치하는 TPR는 각각 약 74.31%에서 76.75%, 76.37%로 증가하였고, TNR는 약 97.82%에서 98.03%, 98.04%로 증가하였다. 반면, 실제 값은 Negative이나 예측 값이 Positive인 FPR은 2.18%에서 1.97%와 1.96%로 감소하였고, 실제 값은 Positive이나 예측 값은 Negative인 FNR은 25.6%에서 23.25%, 23.63%로 감소하였다. 또한, 제로 패딩을 적용한 실험과 테두리 외삽법을 적용한 실험의 비교를 통해, 가상의 픽셀을 가장 자리의 픽셀로 복사해 테두리를 채우는 테두리 외삽법보다 제로 패딩을 적용하여 학습하는 것이, 객체 학습 가중치에 노이즈를 주지 않고 성능을 향상시켰으며, 객체의 종횡비 유지에 따라, FPR과 FNR이 개선되는 것을 확인할 수 있다.

VI. 결론 및 향후 과제

본 논문은 원본의 해상도 종횡비 유지를 통한 객체 비율 보존 기법을 제안한다. 본 논문에서는 패션 영상 이미지 분류 시스템에 객체 비율 보존 기법을 적용한 후, 신경망을 학습하여 성능을 비교하였다. 객체 비율 보존 기법은 원본의 해상도 비를 유지하기 위해 가상의 픽셀을 삽입하는 기법으로 고유한 객체의 비율을 유지시키고 픽셀의 왜곡 현상을 방지한다. 실험에서는 비율 보존 모듈을 적용하지 않은 경우와 비율 보존 모듈을 적용한 경우의 비교를 통해, 제안하는 비율 보존 기법 적용 시 성능 향상을 확인하였으며, ResNet과 DenseNet을 사용한 실험 모두 유사한 성능 향상을 확인하였다. 또한, 제로 패딩을 이용한 기법과 테두리 외삽법과의 비교를 통해, 제로 패딩을 적용하여 학습하는 것이 학습 가중치의 영향을 주지 않으므로 성능 향상에 더 큰 도움이 됨을 보여준다. 오차 행렬 분석을 통해 정답률에 해당하는 평가 지표인 TPR와 TNR를 향상시키고, FPR와 FNR을 감소시켜 클래스의 오분류 또한 개선하는 것을 확인하였다.

이것은 CNN 학습 과정에서 일어나는 픽셀 왜곡 현상을 해결하고, 비율 정보가 중요한 의류 클래스의 고유한 객체의 비율을 유지시킴으로써, 각기 다른 비율을 가지는 의류 분류 시스템에서의 성능 향상을 보여주었다. 제안하는 기

술은 패션 영상 이미지 분류 뿐만 아니라, 고유한 객체의 비율이 중요하게 고려되는 다른 모든 분야의 인식 성능을 개선할 수 있을 것으로 기대된다.

참고 문헌 (References)

- [1] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," *2017 International Conference on Engineering and Technology (ICET)*, Antalya, Turkey, pp. 1-6, 2017.
doi: <https://doi.org/10.1109/ICEngTechnol.2017.8308186>.
- [2] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770-778, 2016.
- [3] Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4700-4708, 2017.
- [4] Islam, Md Amirul, et al. "Position, padding and predictions: A deeper look at position information in cnns." *arXiv preprint arXiv:2101.12322*, 2021.
doi : <https://doi.org/10.48550/arXiv.2101.12322>
- [5] Kayhan, Osman Semih, and Jan C. van Gemert. "On translation invariance in cnns: Convolutional layers can exploit absolute spatial location." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 14274-14285, 2020.
- [6] Li, Debang, Junge Zhang, and Kaiqi Huang. "Learning to learn cropping models for different aspect ratio requirements." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12685-12694, 2020.
- [7] Esmaeili, Seyed A., Bharat Singh, and Larry S. Davis. "Fast-at: Fast automatic thumbnail generation using deep neural networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4622-4630, 2017.
- [8] Hashemi, Mahdi. "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation." *Journal of Big Data* 6.PP. 1-13, 2019.
- [9] Srinivasan, Pratul P., et al. "Pushing the boundaries of view extrapolation with multiplane images." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 175-184, 2019.
- [10] Bailey, Donald G., and Anoop S. Ambikumar. "Border handling for 2D transpose filter structures on an FPGA." *Journal of Imaging* 4.12, 138, 2018.
doi : <https://doi.org/10.3390/jimaging4120138>
- [11] Hou, Yuxin, et al. "Learning Attribute-driven Disentangled Representations for Interactive Fashion Retrieval." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 12147-12157, 2021.
- [12] Chen, Ming, et al. "Improving fashion landmark detection by dual attention feature enhancement." *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. pp. 0-0,

- 2019.
- [13] Ak, Kenan E., et al. "Learning attribute representations with localization for flexible fashion search." Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7708-7717, 2018.
- [14] Laenen, Katrien, Susana Zoghbi, and Marie-Francine Moens. "Cross-modal search for fashion attributes." Proceedings of the KDD 2017 Workshop on Machine Learning Meets Fashion. Vol. 2017. ACM, pp. 1-10. 2017.
doi : 10.475/123_4
- [15] Ak, Kenan E., et al. "Fashionsearchnet: Fashion search with attribute manipulation." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018.
- [16] Kaessli, Nour, Romain Guigoures, and Reza Shirvany. "Sizenet: Weakly supervised learning of visual size and fit in fashion images." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.
- [17] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25, pp.1097-1105. 2012
- [18] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition."arXiv preprint arXiv: 1409.1556, 2014.
doi : <https://doi.org/10.48550/arXiv.1409.1556>
- [19] Susmaga, Robert. "Confusion matrix visualization." Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM '04 Conference held in Zakopane, Poland, May 17-20, 2004. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

저 자 소 개



최 은 지

- 2020년 2월 : 건국대학교 영미어문학 졸업(학사)
- 2022년 2월 : 건국대학교 스마트ICT융합공학과 졸업(석사)
- ORCID : <https://orcid.org/0009-0006-0596-6084>
- 주관심분야 : 스마트미디어시스템, 멀티미디어 검색, 인공지능, 컴퓨터비전



박 경 기

- 1987년 2월 : 연세대학교 전자전산기공학과 졸업(학사)
- 1989년 12월 : University of Michigan, Ann Arbor, 전기전산기공학과 졸업(석사)
- 1999년 5월 : Syracuse University, 전산학과 졸업(박사)
- 1999년 6월 ~ 2003년 8월 : LG전자기술원 책임연구원/그룹장
- 2003년 9월 ~ 현재 : 건국대학교 컴퓨터공학과스마트ICT융합공학과 교수
- 2017년 10월 ~ 현재 : 국립전파연구원 멀티미디어부호화 전문위원회 대표전문위원
- 2019년 10월 ~ 현재 : IEEE-SA Interfacing Cyber and Physical World Working Group 의장
- ORCID : <https://orcid.org/0000-0002-1153-4038>
- 주관심분야 : 스마트미디어시스템, 멀티미디어검색, 영상처리, 멀티미디어/메타데이터 처리, 메타버스