



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

M.S. THESIS

Effect of Rollback in Triplet Loss based Deep Network for Person Re-Identification

사람 재인식을 위한 트리플렛 손실함수 기반의 심층
네트워크에서 롤백의 효과

BY

Sanghun Kim

August 2019

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

M.S. THESIS

Effect of Rollback in Triplet Loss based Deep Network for Person Re-Identification

사람 재인식을 위한 트리플렛 손실함수 기반의 심층
네트워크에서 롤백의 효과

BY

Sanghun Kim

August 2019

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Effect of Rollback in Triplet Loss based Deep Network for Person Re-Identification

사람 재인식을 위한 트리플렛 손실함수 기반의 심층
네트워크에서 롤백의 효과

지도교수 최진영

이 논문을 공학석사 학위논문으로 제출함

2019년 8월

서울대학교 대학원

전기 컴퓨터 공학부

김상훈

김상훈의 공학석사 학위 논문을 인준함

2019년 8월

위원장:	조남익
부위원장:	최진영
위원:	오성희

Abstract

Person re-identification is the technique for matching information of the same person among images taken by several non-overlapping camera. It can be usefully applied to human tracking, visual surveillance, forensics and so on. Person re-identification has two main branches: "Cross-entropy method" based on image classification technique and "Triplet method" using image pair input and triplet loss function. Because each method has its own advantage, it is difficult to compare which method is proper to solve the open problem like person re-identification. In this paper, we compare Cross-entropy method and Triplet method, and apply the effect of rollback, which was previously used only for Cross-entropy method, to Triplet method. In addition, we propose an algorithm that can achieve better performance by applying both Cross-entropy method and Triplet method simultaneously.

keywords: Person Re-Identification, Rollback, Triplet Loss, Image-Classification

student number: 2017-20953

Contents

Abstract	i
Contents	ii
List of Tables	iv
List of Figures	v
1 서론	1
1.1 연구 배경	1
1.2 연구 내용	3
2 배경 지식	4
2.1 선행 연구	4
2.1.1 사람 재인식 데이터셋의 경향	4
2.1.2 사람 재인식 기술의 방법	5
2.1.3 학습 전략 개선에 대한 기존의 연구들	6
2.2 배경 이론	7
2.2.1 이미지 분류를 위한 딥러닝 모델에서 cross-entropy 손실함수	7
2.2.2 사람 재인식을 위한 딥러닝 모델에서의 triplet 손실함수	8
2.2.3 데이터 증가	10

3 제안 방법	12
3.1 Cross-entropy 방법의 사람 재인식 네트워크 구조	12
3.2 Triplet 방법의 사람 재인식 네트워크 구조	13
3.3 Cross-entropy 방법과 Triplet 방법을 통합한 사람 재인식 네트워크 구조	15
3.4 롤백(Rollback)의 적용	17
4 실험 결과	19
4.1 합성곱 신경망 기반 사람 재인식 방법의 학습과 테스트	19
4.1.1 사람 재인식 기술의 성능 측정 메트릭	19
4.1.2 학습 방법	22
4.1.3 테스트 방법	22
4.2 실험 결과	23
4.2.1 실험 세부사항	23
4.2.2 Cross-entropy 방법의 비교 실험	24
4.2.3 Triplet 방법의 비교 실험	25
4.2.4 통합 네트워크의 비교 실험	26
4.2.5 통합 네트워크에서 롤백의 효과	27
4.2.6 실험 결과 종합	28
4.2.7 각 scheme 별 통합네트워크의 성능	29
4.2.8 기존에 가장 높은 성능을 발휘하는 사람 재인식 네트워크들 과 통합 네트워크의 성능 비교	30
5 결론	31
Abstract (In Korean)	37

List of Tables

2.1	대표적인 이미지 기반 사람 재인식 데이터셋, 순서대로 [1-5]	5
4.1	query-gallery 거리 행렬	20
4.2	이미지 분류에 대한 기학습 유무에 따른 사람 재인식 알고리즘의 Market-1501에서 성능 비교	24
4.3	Cross-entropy 방법으로 구현된 사람 재인식 네트워크의 성능과 이 네트워크에 롤백을 적용했을 때 Market-1501에서 성능 비교	25
4.4	기본적인 triplet 손실함수 기반의 사람 재인식 네트워크의 성능과 이 네트워크에 롤백을 적용했을 때 Market-1501에서 성능 비교	26
4.5	통합 네트워크의 Market-1501에서 성능 비교	27
4.6	통합 네트워크의 성능과 이 네트워크에 롤백을 적용했을 때 Market-1501에서 성능 비교	27
4.7	본 논문에서 제시한 네트워크들의 Market-1501에서 성능 비교	28
4.8	통합 네트워크의 scheme별 성능	29
4.9	기존 최고 성능(state-of-the-art)의 사람 재인식 기술과 통합 네트워크의 Market-1501에서 성능 비교	30

List of Figures

2.1	cross-entropy 손실함수 역할 도식화	8
2.2	triplet 손실함수 역할 도식화	9
3.1	이미지 분류 네트워크의 구분	13
3.2	triplet 손실함수를 이용하는 사람 재인식 네트워크	14
3.3	통합된 사람 재인식 네트워크	16
4.1	사람 재인식 과정	20
4.2	Recall-Precision 곡선 및 CMC 곡선 예시	21
4.3	이미지 분류를 위한 딥러닝 네트워크 구조	23

Chapter 1

서론

1.1 연구 배경

사람 재인식 (Person re-identification) 기술은 여러 개의 겹치지 않는 카메라 뷰(camera view)로부터 촬영된 사람들의 정보 중 동일한 사람의 정보를 매칭시키는 기술이다. 이러한 기술들은 대인 이동 경로 추적, 감시 시스템(visual surveillance), 포렌식(forensic) 등에 유용하게 적용될 수 있다. 사람 재인식 기술은 여러 이미지 중 동일한 사람의 이미지를 분류해야 한다는 점에서 이미지 분류(image classification) 기술과 깊은 연관이 있다. 과거에 비해 이미지 분류 기술은 눈부신 성능 향상을 이루었는데, 빅데이터(big data)와 딥러닝(deep learning) 알고리즘이 큰 역할을 했다 [6]. 딥러닝을 통한 이미지 분류 기술의 성능 향상에 힘 입어, 사람 재인식 알고리즘에 기학습된(pre-trained) 이미지 분류 딥러닝 네트워크가 사용되면서 괄목할만 한 성능향상을 이루었다 [7,8]. 하지만 이미지 분류 문제는 학습 과정에서의 클래스가 평가 과정의 클래스와 동일한 닫힌 집합 문제 [9]인 반면 사람 재인식 문제는 학습 과정에서의 클래스와 평가 과정의 클래스가 다르다. 이에 따라, 사람 재인식 문제는 이미지 분류 문제에 비해 일반화(generalization) 성능을 확보하기가 어렵다.

열린 집합 문제(open set problem)를 풀기 위한 고전적인 방법은 기준(anchor) 이미지, 동일한 사람(positive)의 이미지 그리고 다른 사람(negative)의 이미지 튜플을 입력으로 하여 triplet 손실함수를 사용하는 metric learning 방법 [10-12]이다. 이는

사람 재인식 기술과 같은 열린 집합 문제에서 객체들의 신원(identity) 정보가 아닌 객체간의 거리정보를 이용하기 때문에 적합한 방법이다. 그러나 [13]에서, 이미지 분류와 같이 주로 닫힌 집합 문제에서 사용되었던 cross-entropy 손실함수도 사람 재인식 문제에서 사용될 수 있음을 검증하였고, triplet 손실함수를 사용하는 것보다 더 좋은 성능을 얻을 수 있다고 보고하였다. 한편, 이 후 [14]에서는 triplet 손실함수를 이용한 사람 재인식 네트워크에서 입력으로 기준 이미지와 가장 거리가 가까운 동일한 사람·다른 사람의 이미지를 의미하는 hard sample을 선택하여 학습할때 cross-entropy 손실함수를 활용하는 사람 재인식 방법보다 더 좋은 성능을 얻을 수 있다고 보고하였다. 각각의 방법들의 고유한 특성때문에 cross-entropy 손실함수 기반의 방법과 triplet 손실함수 기반의 방법 중 어떤 방법이 사람 재인식과 같은 열린 집합 문제를 해결하는데 더 유력한 방법인지 결정하기는 어렵다.

본 논문에서는 기존의 cross-entropy 손실함수 기반의 사람 재인식 기술과 triplet 손실함수 기반의 사람 재인식 기술을 소개하고, 각 방법론을 효과적이고 동등하게 비교하기 위하여 추가적인 확장(Add-on)을 사용하지 않고 단지 학습 방법을 개선하는 롤백(Rollback) 방법에 대한 유용성을 검토하였다. 롤백 방법은 cross-entropy 손실함수 기반의 사람 재인식 방법에 대해서만 검증이 되었고 triplet 손실함수 기반의 방법에 대해서 검증되지는 못하였다. 본 논문에서는 여러가지 실험에 기반한 검증을 통해 cross-entropy 손실함수 기반의 사람 재인식과 triplet 손실함수 기반의 사람 재인식 기술의 장단점을 비교하여 성능을 측정하였으며, 더 나아가 두 방법을 동시에 적용하는 알고리즘을 제안하였다.

1.2 연구 내용

본 논문에서는 cross-entropy 손실함수 기반의 딥러닝 모델을 이용한 롤백 방법 [15]을 분석함과 동시에 사람 재인식 기술의 또 다른 방향인 triplet 손실함수를 이용하는 방법에 롤백을 적용함으로써 성능 향상을 이뤄 낼 수 있음을 보였다. 나아가 사람 재인식 기술의 두가지 방향성을 하나의 네트워크로 통합하는 방법을 제안하였고 롤백을 적용할 경우 높은 성능을 보일 수 있음을 보였다. 이 논문의 요지는 다음과 같다.

- cross-entropy 손실함수 기반, triplet 손실함수 기반의 사람 재인식 알고리즘을 분석하고 비교한다.
- triplet 손실 함수를 이용하는 사람 재인식 알고리즘에 롤백을 적용하면 정확도를 높일 수 있음을 보인다.
- 사람 재인식 기술의 두가지 방향성을 하나로 통합하는 네트워크를 제시하고 해당 네트워크에 롤백을 적용함으로써 기존 사람 재인식 분야에서 state-of-the-art 성능을 보였던 cross-entropy 손실함수 기반의 사람 재인식 기술에 롤백을 적용한 것 [15]보다 더 높은 정확도를 발휘함을 보인다.

Chapter 2

배경 지식

2.1 선행 연구

2.1.1 사람 재인식 데이터셋의 경향

Table 2.1에서 이미지 기반 사람 재인식분야에서 활용되는 대표적인 데이터셋을 정리하였다. 비교적 최근에 나온 데이터셋을 왼쪽부터 나열하였는데 최근의 데이터셋은 다음과 같은 특성을 보인다.

- 데이터셋의 크기가 커졌다.
- 데이터셋을 촬영하는데 많은 수의 카메라가 사용됐다.

사람 재인식 데이터셋은 여러 카메라에서 촬영된 사람들의 이미지에 대하여 사람 검출(person detection)을 수행하고 동일한 사람에 대해 같은 신원(identity)을 부여함으로써 만들어진다. 하지만 사람 재인식 데이터셋은 그 크기가 커질수록 사람이 직접 검출하기가 어렵기 때문에, 검출기(detector) 사용이 불가피한데 이 역시도 완전한 검출이 불가능하므로 검출기를 사용하더라도 사람이 보조적으로 분류 작업을 수행해야만한다. 이러한 특징은 사람 재인식 데이터셋의 크기를 늘리는데 한계점으로 작용한다. 그럼에도 초기의 사람 재인식 데이터셋의 크기에 비해 최근의 데이터셋의 크기는 큰 폭으로 증가하는 추세이다. 예를 들어, 2007년에 공개된 VIPeR에 비해 2017년에 공개된 Duke의 이미지 데이터의 갯수는 약 26배 정도 증가하였으며,

PRID를 제외하고는 시간이 흐름에 따라 데이터셋의 크기는 큰 폭으로 증가하였다. 이는 사람 재인식 딥러닝 모델을 연구하는데 있어 많은 데이터가 필요하다는 경향을 보여주며, 앞으로도 데이터셋의 크기는 계속해서 증가할 것임을 예측할 수 있다. 추가로, 데이터셋을 수집하는데 사용되는 카메라의 갯수가 점점 증가하고 있는데 Duke의 경우엔 카메라가 8개까지 사용되었다. 카메라 갯수가 많아진다는 것은 좋은 일반화 성능을 내기 위해서 다양한 각도에서 취득한 이미지에 대해서도 정확하고 강인한 판단을 내릴수 있어야함을 의미한다.

데이터셋	Duke	Market	CUHK01	PRID	VIPeR
신원(ID)의 수	1,812	1,501	971	200	632
이미지 개수	36,441	32,668	3,884	1,134	1,264
카메라 개수	8	6	2	2	2
공개 연도	2017	2015	2012	2011	2007
레이블링 방법	hand/DPM	hand/DPM	hand	hand	hand

Table 2.1: 대표적인 이미지 기반 사람 재인식 데이터셋, 순서대로 [1-5]

2.1.2 사람 재인식 기술의 방법

사람 재인식 기술은 크게 두 가지 지도학습(supervised learning) 방법으로 구현된다. 하나는 cross-entropy 손실함수에 기반한 방법이며, 다른 하나는 triplet 손실함수에 기반한 방법이다. 본 논문에서는 cross-entropy 손실함수에 기반한 방법을 "Cross-entropy 방법", triplet 손실함수에 기반한 방법을 "Triplet 방법"이라 칭한다. 두 방법의 세부적인 작동 과정은 4.1.2 섹션에서 다루며 개요는 아래와 같다.

Cross-entropy 방법: 먼저, 사람 재인식 기술은 충분한 양의 데이터셋 확보가 어려우므로 ImageNet과 같은 풍부한 데이터에서 기학습된 모델(pre-trained model)을 기초상태(initial state)로 하여 세밀조정(fine-tuning) 하는 식으로 학습이 진행 된다. [2]에서 구현된 Cross-entropy 방법은 이미지 분류를 목적으로 구현된 심층 네트워크

크 [16]에서 사람 재인식 데이터셋의 신원 수만큼 완전 연결층(fully-connected layer)의 출력을 조정한 뒤 세밀 조정을 수행하여 이뤄진다. 이 경우, 이미지 분류를 위한 심층 네트워크에 내재된 특징점을 그대로 활용하면서 사람 재인식 데이터셋의 특징점을 추가로 학습할 수 있으며, 이로 인해 **2.1.1 섹션**에서 기술한 바와 같이 데이터셋의 크기가 크지 않은 사람 재인식 기술에 유용하게 활용된다.

Triplet 방법: 이는 이미지 쌍의 입력으로부터 특징 벡터(feature vector)를 얻어내고, 얻어진 특징 지도와 적당한 거리 메트릭을 통한 triplet 손실 함수의 크기를 줄이는 모델을 활용하는 방법이다. 이 방법은 네트워크의 학습과정에 있어 학습 데이터의 신원 정보를 직접적으로 활용하지 않기 때문에 사람 재인식과 같은 열린 집합문제에 적합하다.

2.1.3 학습 전략 개선에 대한 기존의 연구들

사람 재인식 분야에서 학습 전략 개선을 통해 성능을 높이려는 연구가 진행되고 있다. [17]에서는 triplet 손실함수가 테스트 데이터에 대해 일반화 성능이 좋지 못하다고 주장하였고, 이를 개선하기 위해 quadruplet 손실함수를 제안하였다.

[18]에서는 모델 증류(model distillation)를 이용하여 높은 성능을 보이지만 계산 복잡도가 높은, 교사 네트워크(teacher network)로부터 상대적으로 계산복잡도가 낮은 학생 네트워크(student network)가 학습할 수 있도록 하였고, 이러한 학생 네트워크들을 앙상블(ensemble)하여 성능을 측정하면 교사 네트워크보다 높은 성능을 달성할 수 있다고 주장하였다. [19]에서는 SVDNet(Singular Vector Decomposition Network)를 제안하였는데, 이는 딥러닝 네트워크의 완전연결층(Fully-connected layer)의 가중치가 서로 연관(correlate)돼 있는 문제를 해결하기 위해 특이점 분해(Singular vector decomposition) 방법을 도입했다.

[15]에 의하면 이미지 분류를 위한 합성곱 신경망(convolutional neural network) 기반 딥러닝 모델을 이용해 사람 재인식 기술을 구현 할 때 데이터셋이 부족한 문제를 해결하기 위해 기학습된 네트워크로부터 세밀 조정하는 방법을 이용하는데, 이때 역전파(backpropagation) 과정에서, high-level layer의 과적합(overfitting)에 의해,

low-level layer에 대한 그래디언트(gradient)의 영향력이 줄어들어 해당층이 충분히 세밀 조정되지 못한다. 이 때, low-level layer를 반복해서 학습시킬 수 있는 롤백(rollback)을 통해 사람 재인식 기술의 성능을 높일수 있다. 본 논문에서는 이 방법의 다양한 형태의 적용성에 대해 검토하였으며, 이 방법의 자세한 과정은 3.4 섹션에서 다룬다.

2.2 배경 이론

2.2.1 이미지 분류를 위한 딥러닝 모델에서 cross-entropy 손실함수

합성곱 신경망 기반의 이미지 분류는 다양한 기술들이 있다 [6, 16, 20, 21]. 이러한 기술들은 네트워크의 세부적인 구현에 차이가 있을 뿐 작동의 전반적인 틀은 동일하다. **Figure 2.1**과 같이 cross-entropy 손실함수를 통해, 입력 이미지가 네트워크를 통과했을 때의 결과인 특징벡터가 해당 이미지의 클래스를 one-hot형태로 나타낸 벡터와 유사해지도록 한다. 여기서 cross-entropy 손실함수는 다음과 같이 계산된다.

먼저 학습 데이터 샘플을 $\{\mathbf{x}_i\}_{i=1}^N$ 이라 하자. 이 때 N 은 총 학습데이터의 개수가 된다. 이미지 분류 데이터셋은 데이터 샘플마다 구분하려는 클래스 개수(C)를 차원으로 가지는 one-hot 벡터 형태의 레이블이 존재하는데 이를 $\{\mathbf{y}_i\}_{i=1}^N$ 이라 하자. 그러면 각각의 입력 이미지 \mathbf{x}_i 가 이미지 분류 네트워크를 통과하고 완전연결층을 지나면 C 차원의 벡터를 생성해내는데 이 과정을 $\mathcal{P}(\mathbf{x}_i|\theta_1, \dots, \theta_n)$ 로 표현하면 cross-entropy 손실함수 $\mathcal{L}_{entropy}$ 은 다음과 같다. (여기서 $\theta_1, \dots, \theta_n$ 는 딥러닝 네트워크 각 층의 가중치 파라미터를 나타낸다.)

$$\mathcal{L}_{entropy}(\mathbf{x}_i, \mathbf{y}_i) = -\frac{1}{N} \sum_{i=1}^N \mathbf{y}_i^T \log \mathcal{P}(\mathbf{x}_i|\theta_1, \dots, \theta_n)$$

위 손실함수는 예측값($\mathcal{P}(\mathbf{x}_i|\theta_1, \dots, \theta_n)$)과 실측값(\mathbf{y}_i)의 차이가 클수록 결과 값이 커지고 반대의 경우엔 작아지므로 제대로 동작한다.

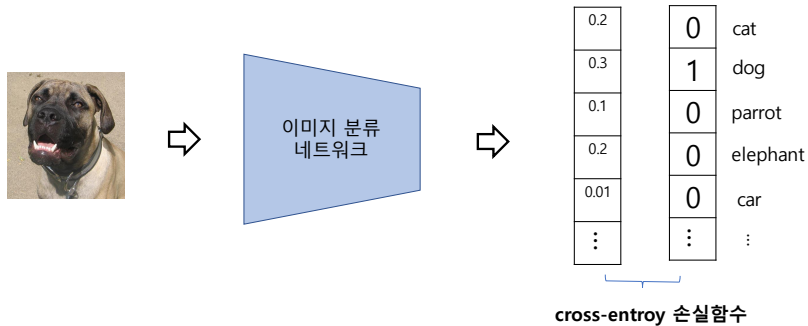


Figure 2.1: cross-entropy 손실함수 역할 도식화

2.2.2 사람 재인식을 위한 딥러닝 모델에서의 triplet 손실함수

[10]에서는 사람의 얼굴 정보를 포함하고 있는 데이터셋으로부터 동일한 사람인지 아닌지 구별해내기 위해 FaceNet이라는 딥러닝 네트워크를 활용하였다. 여기서 triplet 손실 함수를 활용함으로써, 동일한 얼굴의 이미지 쌍에 대해서는 특징 벡터 (feature vector) 사이의 거리를 줄이고 반대의 경우엔 거리를 늘리도록 네트워크를 학습시켰다. 이 때, FaceNet의 판별 대상이었던 사람의 얼굴에서 몸 전체로 도메인을 확장시켜 사람 재인식 기술에도 적용할 수 있다. 이러한 네트워크에 적용되는 triplet 손실함수의 개념은 **Figure 2.2**와 같이 나타낼 수 있다.

사람 재인식 네트워크가 triplet 손실함수를 통해 학습을하게 되면 기준이 되는 이미지의 특징 벡터와 동일한 사람의 이미지(positive)의 특징 벡터 사이의 거리는 가까워지고, 다른 사람의 이미지(negative)의 특징 벡터 사이의 거리는 멀어지게 된다. 기준이 되는 사람의 이미지(anchor), 동일한 사람의 이미지(positive), 다른 사람의 이미지(negative)를 각각 x^A, x^+, x^- 라 하자. 그리고 입력 이미지 x 가 사람 재인식 네트워크를 통과한 뒤 정규화된(normalized) 결과를 $\rho(x)$ 라 하면, 동일한 사람의 이

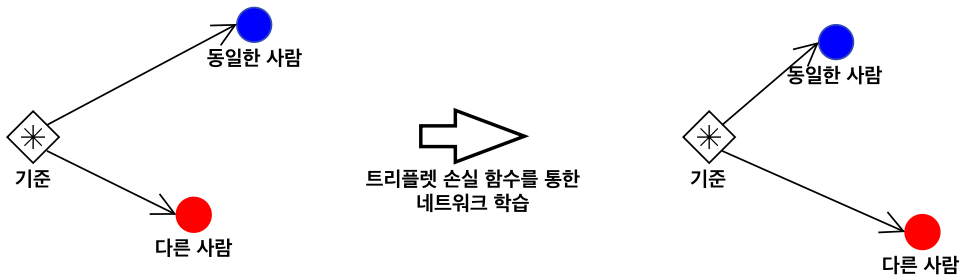


Figure 2.2: triplet 손실함수 역할 도식화

미지 사이의 거리가 다른 사람의 이미지 사이의 거리보다 가깝게 하기 위해서 다음과 같은 수식을 세울 수 있다.

$$\|\rho(x^A) - \rho(x^+)\| + \alpha < \|\rho(x^A) - \rho(x^-)\|.$$

여기서 α 는 동일한 사람과 다른 사람을 구별하는 거리의 margin을 의미한다. 여기서, 손실 함수의 형태로 위 수식을 재구성하면 triplet 손실 함수 \mathcal{L}_{tri} 은 다음과 같이 표현할 수 있다.

$$\mathcal{L}_{tri} = \sum \max(0, \|\rho(x^A) - \rho(x^+)\| - \|\rho(x^A) - \rho(x^-)\| + \alpha).$$

추가로 노름(norm)의 크기가 과도하게 커지지 않도록 정규항(regularizer) \mathcal{L}_{reg} 을 다음과 같이 정의하면,

$$\mathcal{L}_{reg} = \|\rho(x^A) - \rho(x^+)\|^2 + \|\rho(x^A) - \rho(x^-)\|^2$$

와 같이 된다. 최종 손실함수 $\mathcal{L}_{triplet}$ 은 다음과 같이 표현이 가능하다.

$$\mathcal{L}_{triplet} = \mathcal{L}_{tri} + \mathcal{L}_{reg}.$$

2.2.3 데이터 증가

사람 재인식 분야에서 데이터가 부족한 문제를 해소하기 위한 수단으로 데이터 증가(data augmentation) 기법이 활용된다 [22–25]. 다양한 데이터 증가 기법이 있지만 본 논문의 실험에서는 임의 제거(random erasing) 기법 [23]을 사용한다. 임의 제거 기법은 학습데이터에 대해 임의의 교합(occlusion)을 생성하여 학습데이터의 부족으로 인하여 네트워크가 과적합 되는 것을 방지하고 네트워크를 강인하게 학습할 수 있게 한다.

임의 제거의 전체 과정은 **Algorithm 1**과 같다. 먼저, 0과 1사이의 임의의 값이 할당되는 p_1 이 p 보다 작을 때 임의 제거가 수행된다. 간단히 말하면 p 의 확률로 임의 제거가 수행된다. 이 때, p 는 사용자가 직접 설정하는 하이퍼파라미터다. mini-batch 상의 이미지 I 에 대하여, I 의 크기를 $S = W \times H$ 라 하자. 하이퍼파라미터 s_l, s_h, r_1, r_2 에 대해 임의 제거될 직사각형 영역의 넓이를 S_e 라 하면, $\frac{S_e}{S}$ 는 구간 $[s_l, s_h]$ 내에 존재한다. 또, 임의 제거할 직사각형 영역의 종횡비를 r_e 라 하면, 이는 구간 $[r_1, r_2]$ 내에 존재한다. 그 결과 임의 제거할 직사각형 영역 I_e 의 크기는 높이를 $H_e = \sqrt{S_e \times r_e}$, 너비를 $W_e = \sqrt{\frac{S_e}{r_e}}$ 로 표현할 수 있다. 그러면, I 내부의 임의의 점 $\mathcal{P} = (x_e, y_e)$ 를 정하여, 직사각형 형태의 임의제거할 영역 $I_e = (x_e, y_e, x_e + W_e, y_e + H_e)$ 를 선택할 수 있다(제거영역의 좌상단, 우하단 좌표를 표현). 여기서 $x_e + W_e > W$, $y_e + H_e > H$ 인 경우, 적절한 영역 선택이 되지 않았으므로 영역을 다시 선택하고, 적절한 영역이 선택될 때까지 반복한다. 적절한 영역이 선택된 경우 해당 영역의 픽셀 값을 모두 임의의 값으로 설정하면 임의 제거 과정이 끝난다.

Algorithm 1: 임의 제거 기법

Input : I : 입력 이미지,

W, H : 이미지 너비, 높이,

S : 이미지 면적,

p : 제거 확률,

s_l, s_h : 제거할 영역의 넓이 비율(최소, 최대),

r_1, r_2 : 제거할 영역의 중횡비(최소, 최대).

Output: 임의 제거된 이미지 I^*

Initialize $p_1 \leftarrow 0$ 과 1사이의 임의의 값

if $p_1 \geq p$ **then**

$I^* \leftarrow I$

return I^*

else

while True **do**

$S_e \leftarrow \text{Rand}(s_l, s_h) \times S$

$r_e \leftarrow \text{Rand}(r_1, r_2)$

$H_e \leftarrow \sqrt{S_e \times r_e}, W_e \leftarrow \sqrt{\frac{S_e}{r_e}}$

$x_e \leftarrow \text{Rand}(0, W), y_e \leftarrow \text{Rand}(0, H)$

if $x_e + W_e \leq W$ and $y_e + H_e \leq H$ **then**

$I_e \leftarrow (x_e, y_e, x_e + W_e, y_e + H_e)$

$I(I_e) \leq \text{Rand}(0, 255)$

$I^* \leftarrow I$

return I^*

end if

end while

end if

Chapter 3

제안 방법

3.1 Cross-entropy 방법의 사람 재인식 네트워크 구조

본 연구에서 활용된 Cross-entropy 방법의 사람 재인식 네트워크는 ImageNet에서 기 학습된 ResNet-50 [16]을 사용하였다. ResNet-50은 **Figure 3.1**와 같이 특징점 추출기(feature extractor)와 분류기(classifier)로 나눌수 있다. 특징점 추출기는 50개의 합성곱 층으로 이루어져 있고 각각의 합성곱 층 뒤에 배치 정규화 층(batch normalization layer)과 rectified linear unit(ReLU)가 순차적으로 배치된다. ImageNet에서 기 학습된 네트워크는 1000개의 클래스로 분류하도록 설계 되어있기 때문에, 사람 재인식을 위한 네트워크 설계를 위해서는 본 논문에서 사용되는 사람 재인식 데이터셋 Market-1501 [2]의 신원 개수인 751개의 클래스로 분류하는 분류기(classifier)를 새롭게 정의해야 한다. 분류기는 완전연결층(fully connected layer) 두개로 구성되어 있고 특징점 추출기의 마지막 2048차원의 특징 벡터를 입력으로 하고 중간에 512 차원의 병목층(bottleneck layer)을 지나 사람 재인식 학습데이터의 신원 수와 일치하는 751차원의 벡터를 출력하도록 설계되었다.

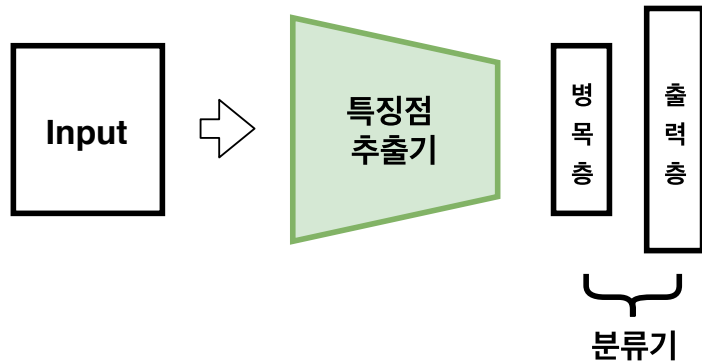


Figure 3.1: 이미지 분류 네트워크의 구분

3.2 Triplet 방법의 사람 재인식 네트워크 구조

Triplet 방법의 사람 재인식 네트워크의 구조는 **Figure 3.2** 와 같이 나타낼 수 있다. conv1부터 Res4_x까지는 ImageNet [26] 데이터셋을 통해 기학습된 ResNet-50 [16] 네트워크를 기본으로 사람 재인식 네트워크를 학습시켰다. Market-1501 학습 데이터에서 기준, 동일한 사람, 다른 사람의 이미지를 샘플링 후 하나의 쌍으로하여 ResNet-50을 통과시키면 각각의 특징 지도를 얻어 낼 수 있는데, 이 특징 지도 각각을 채널별(channelwise)로 전역 평균 풀링(global average pooling)을 수행하면 특징 벡터들을 얻어낼 수 있다. 기준, 동일한 사람, 다른 사람의 이미지로부터 얻어낸 특징 벡터를 이용해 triplet 손실함수를 계산하고, 확률적 그래디언트 하강(Stochastic Gradient Descent) [27]을 이용해 네트워크를 학습시키면 기준 이미지와 다른 사람의 이미지 사이의 거리가 동일한 사람의 이미지 사이의 거리보다 멀어지게 하는 네트워크를 만들어낼 수 있다.

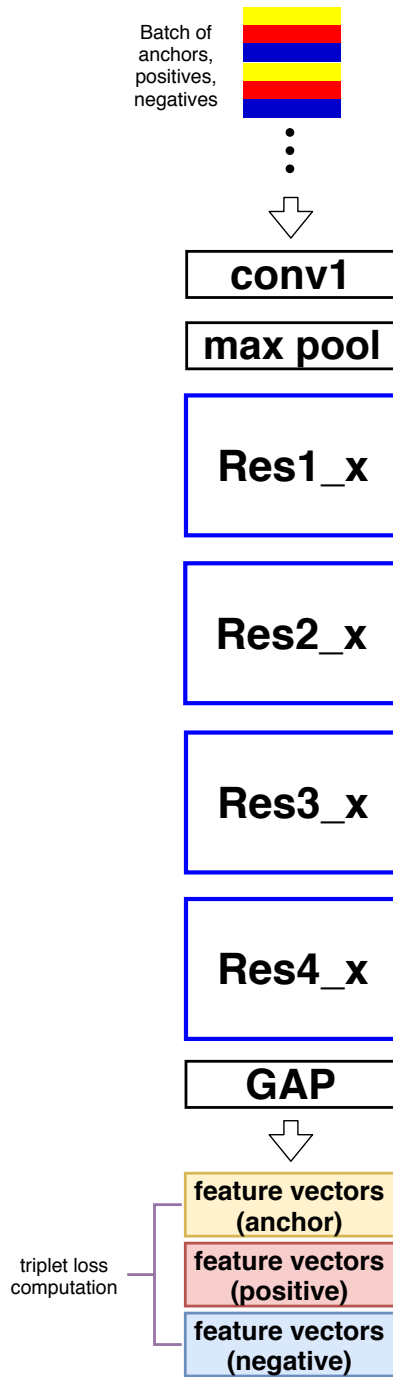


Figure 3.2: triplet 손실함수를 이용하는 사람 재인식 네트워크

3.3 Cross-entropy 방법과 Triplet 방법을 통합한 사람 재인식 네트워크 구조

본 논문에서는 사람 재인식 기술의 두가지 방향성인 Cross-entropy 방법과 Triplet 방법을 통합한 네트워크를 제시하였고 이를 "통합 네트워크(Integrated Network)"라 칭한다. 이 구조에서는 triplet 손실함수를 사용하면서 동시에 cross-entropy 손실함수도 사용한다. 먼저, **Figure 3.2**의 과정을 기본으로 triplet 손실함수 $\mathcal{L}_{triplet}$ 을 계산함과 동시에 광역 평균 풀링(GAP) 층을 지난 특징 벡터를 병목층을 통과하게 하여 512차원으로 만든 후 완전연결층을 통과하게 하여 751개의 차원(Market-1501 데이터셋의 신원 개수)으로 만든다. 기준, 동일한 사람, 다른 사람의 구분 없이 데이터셋 내에 포함된 레이블 정보를 통해 cross-entropy 손실함수 $\mathcal{L}_{entropy}$ 도 함께 계산한다. 그러면 통합 네트워크의 손실함수 $\mathcal{L}_{integrated}$ 는 다음과 같다.

$$\mathcal{L}_{integrated} = \mathcal{L}_{triplet} + \beta \mathcal{L}_{entropy}.$$

여기서 β 는 가중치를 의미한다. 이러한 방법으로 계산되는 손실함수를 통해 역전파를 수행하여 네트워크를 학습시킨다. 이를 그림으로 도식화하면 **Figure 3.3**와 같다.

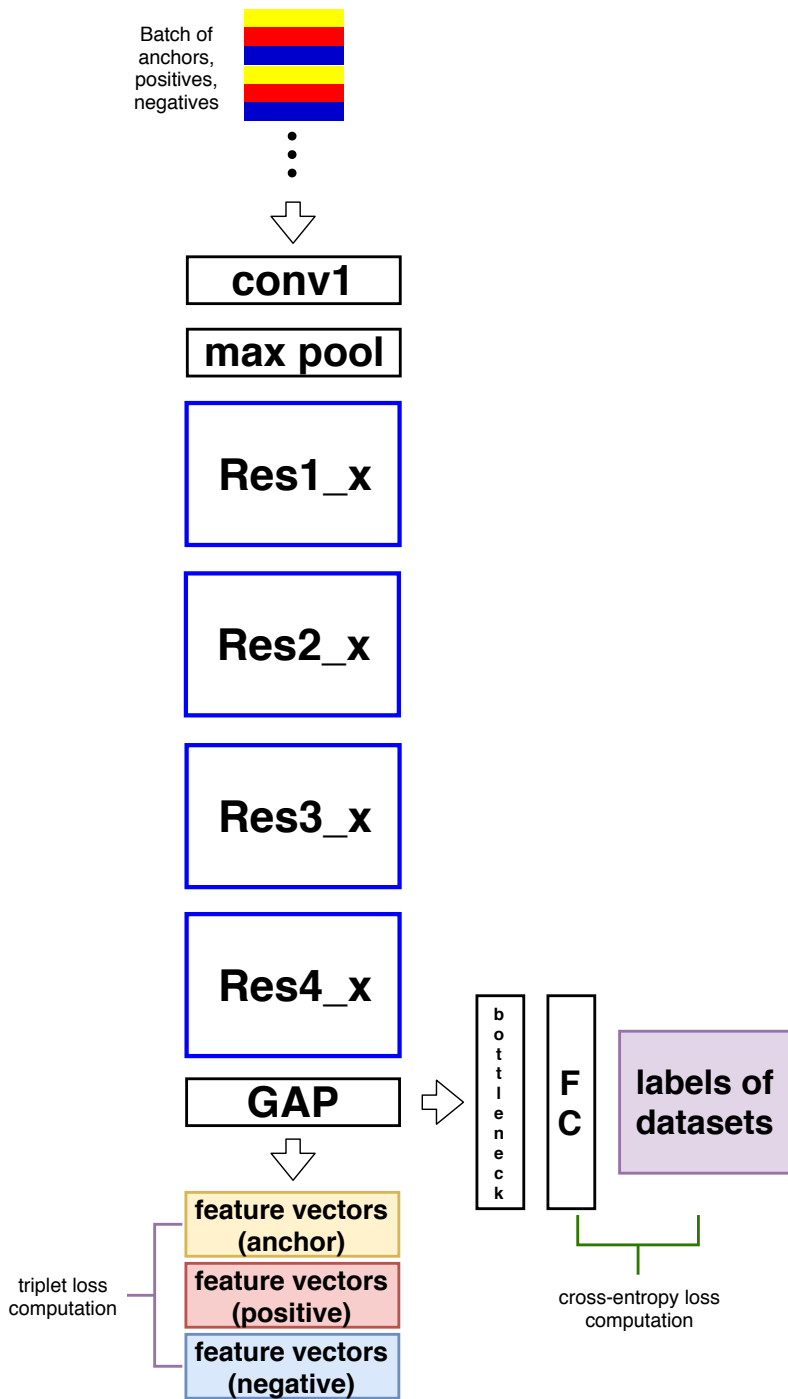


Figure 3.3: 통합된 사람 재인식 네트워크

3.4 롤백(Rollback)의 적용

[15]에 의하면 이미지 분류 목적으로 학습된 합성곱 신경망을 기반으로 사람 재인식을 수행하도록 세밀 조정하는 과정에서, 네트워크의 high-level layer들의 가중치(weight)를 원래 이미지 분류 목적 합성곱 신경망 상태로 돌림으로써(rollback) low-level layer들이 충분히 학습 될 수 있다. 딥러닝 모델의 학습이 진행되면서 high-level layer들이 데이터셋에 과적합(overfitting)되는데, 이 경우 low-level layer들은 손실 함수의 영향이 충분히 미치지 못하게 된다. 롤백을 이용하면 low-level layer들이 충분히 학습될 수 있고, 사람 재인식 데이터셋에 대해서 성능 향상을 보인다. 사람 재인식 데이터셋을 이용한 합성곱 신경망의 롤백 과정을 자세히 명시하면 **Algorithm 2**과 같다.

먼저 기학습된(pre-trained) 이미지 분류 네트워크에서 시작한다. 이 네트워크의 가중치는 완전연결층을 제외한 N 개의 네트워크의 블록에 대해 $\{\theta_1^{(0)}, \dots, \theta_N^{(0)}\}$ 로 표현된다. θ 의 위첨자의 소괄호내 숫자는 네트워크가 사람 재인식 데이터셋에 대해 세밀 조정된 횟수를 의미한다. 이 상태에서 사람 재인식 데이터셋을 이용해 네트워크를 학습시킨다. 이 때의 네트워크 블록들의 가중치는 $\{\theta_1^{(1)}, \dots, \theta_N^{(1)}\}$ 으로 표현 될 수 있다. 이 후, $\theta_1^{(1)}$ 을 제외한 모든 블록의 가중치 $\{\theta_i^{(1)}\}_{i=2}^N$ 를 각각 처음 네트워크의 가중치 $\{\theta_i^{(0)}\}_{i=2}^N$ 로 대체한다. 이 상태에서 사람 재인식 데이터셋을 이용해 다시 한번 네트워크를 학습시키고 $\theta_1^{(2)}, \theta_2^{(2)}$ 를 제외한 모든 블록 $\{\theta_i^{(2)}\}_{i=3}^N$ 을 $\{\theta_i^{(0)}\}_{i=3}^N$ 로 대체한다. 이러한 과정을 네트워크 상의 반복학습이 필요한 가장 마지막 블록이 포함될 때까지 지속한다. 사람 재인식 네트워크를 학습시키는데 있어 롤백을 적용하게 되면 네트워크의 추가적인 확장(add-on) 없이 기본적인 학습 방법 개선으로 성능을 높일 수 있다는 장점이 있다.

Algorithm 2: 사람 재인식 네트워크의 롤백

Parameter: N : 전체 블록의 수, M : (반복적 학습이 필요한) 초기 블록의 수

Parameter: $\theta_1^{(0)}, \dots, \theta_N^{(0)}$: 기학습 합성곱 신경망의 가중치

Input : $\theta_1, \dots, \theta_N$, 데이터셋: (X, Y)

$$\theta_i^{(1)} = \theta_i^{(0)}, \forall i = 1, \dots, N$$

$\hat{\theta}_1^{(1)}, \dots, \hat{\theta}_N^{(1)}$ (사람 재인식 데이터셋에 대하여 세밀 조정)

for $p = 2$ to M **do**

$$\theta_i^{(p)} = \begin{cases} \hat{\theta}_i^{(p-1)} & \text{if } i < p \\ \theta_i^{(0)} & \text{if } i \geq p. \end{cases}$$

$\hat{\theta}_1^{(p)}, \dots, \hat{\theta}_N^{(p)}$ (사람 재인식 데이터셋에 대하여 세밀 조정)

end for

Chapter 4

실험 결과

4.1 합성곱 신경망 기반 사람 재인식 방법의 학습과 테스트

4.1.1 사람 재인식 기술의 성능 측정 메트릭

사람 재인식 알고리즘은 **Figure 4.1**과 같은 과정으로 동작한다. 찾고자 하는 대상을 포함하는 query 이미지가 주어졌을 때, 카메라에 찍힌 신원들의 이미지 모음인 gallery 이미지 내에서 query 이미지와 동일한 사람들의 이미지를 찾아낸다. 이 때, 테스트 과정에서 성능을 측정하기 위해서는 먼저 어떤 query 이미지에 대하여 데이터셋 내에 존재하는 모든 gallery 이미지와의 거리를 전부 구한다. 그 후, 거리가 가까운 gallery 이미지부터 순서대로 나열하는데, gallery 이미지 중 query 이미지와 동일한 사람의 이미지가 앞쪽에 많이 배치 될수록 더 높은 성능을 보인다고 할 수 있다. 이러한 성능 측정을 위한 준비 과정으로 **Table 4.1**과 같이 query 이미지와 gallery 이미지 사이의 거리 정보를 포함하는 행렬을 먼저 만들고 거리가 가까운 순서로 왼쪽에 배치한다. 여기서 $d(.,.)$ 은 두 인자 사이의 $L2$ 거리를 의미하고 사람 재인식 테스트 데이터셋내 query image들을 $\{\mathbf{q}_i\}_{i=1}^n$, gallery 이미지들을 $\{\mathbf{g}_i\}_{i=1}^n$ 과 같이 표현하였다.

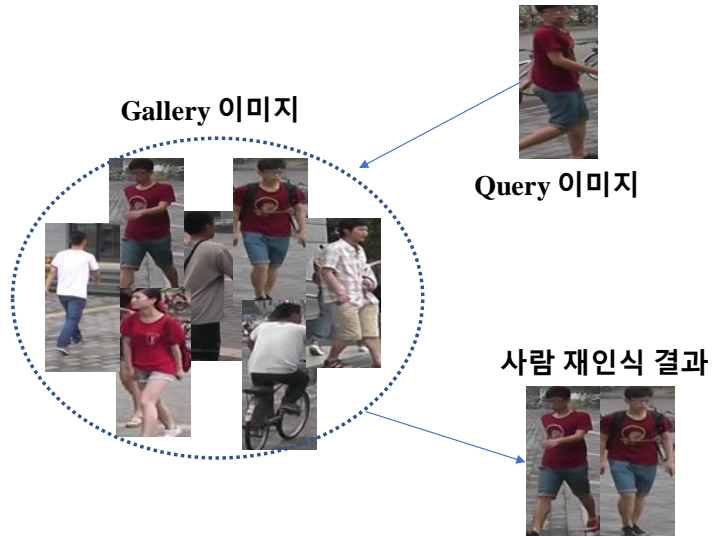
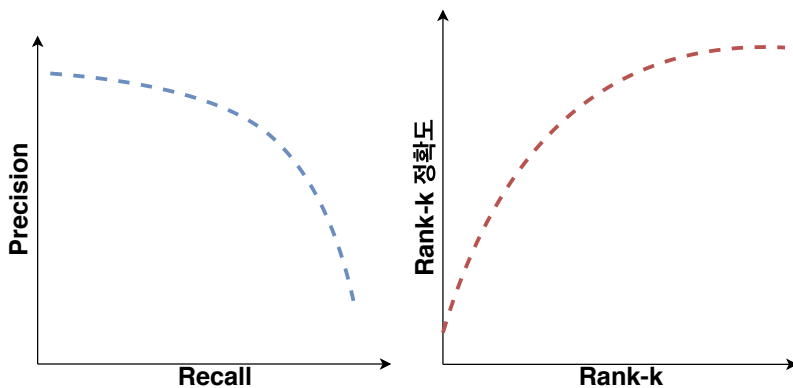


Figure 4.1: 사람 재인식 과정

Table 4.1: query-gallery 거리 행렬

		Gallery image(m개)			
Query image(n개)	$d(q_1, g_1)$	$d(q_1, g_2)$	$d(q_1, g_3)$...	← 오름차순 정렬 수행
	$d(q_2, g_1)$	$d(q_2, g_2)$	$d(q_2, g_3)$...	← 오름차순 정렬 수행
	$d(q_3, g_1)$	$d(q_3, g_2)$	$d(q_3, g_3)$...	← 오름차순 정렬 수행
	⋮	⋮	⋮	⋱	⋮



(a) Recall-Precision 곡선 예시 (b) CMC 곡선 예시

Figure 4.2: Recall-Precision 곡선 및 CMC 곡선 예시

mAP: 사람 재인식 기술의 성능을 측정하는 메트릭 중 하나로 mean Average Precision(mAP)이 있다. 이는 사람 재인식 데이터셋에 존재하는 각각의 query 이미지에 대한 평균 정확도(Average Precision)를 구한 뒤 그 값들의 평균 값을 구하게 된다. 어떤 query 이미지에 대한 평균 정확도를 구하려면, 해당 query 이미지의 사람 재인식 결과에 대한 **Figure 4.2** 형태의 Recall-Precision 곡선을 구해서 곡선 아래 넓이를 구하면 된다. Recall-Precision 곡선이 얻어지는 방식을 **Table 4.1**을 이용해 나타낼 수 있다. 예를 들어, query 이미지 q_1 의 Recall-Precision은 행렬의 첫 번째 행을 순차적으로 한개씩 탐색해가면서 전체 탐색된 gallery 이미지 수에 대한, query 이미지내 사람과 gallery 이미지내 사람의 신원이 일치하는 gallery 이미지 수의 비율이다. 탐색 개수에 따른 Recall-Precision을 그래프로 나타내면 **Figure 4.2**와 같은 경향을 볼 수 있다. 이 때, 그래프 곡선 아랫부분의 넓이를 구한 것이 평균 정확도이고 모든 query의 평균 정확도들의 평균을 구한 것이 mean Average Precision이다.

CMC: 또 다른 사람 재인식 기술 성능 측정 메트릭으로는 Cumulative Matching Characteristic(CMC)가 있다. 이를 이해하기 위해서는 먼저 rank-k 정확도에 대한 이해가 필요한데, 전체 query 이미지에 대해 거리가 가까운 순서대로 gallery 이미지를 나열했을 때(**Table 4.1**) 가깝다고 판단된 k개의 gallery 이미지내의 사람 중에 query 이미지 내의 사람과 신원이 일치하는 gallery가 최소한 하나이상 존재하는 query 이

미지의 비율이다. 예를 들어, **Table 4.1**의 첫 번째 열에서 query 이미지내 사람과 gallery 이미지내 사람의 신원이 일치하는 경우가 5가지라면, rank-1 정확도는 $\frac{5}{n}$ 가 된다. **Figure 4.2**에 CMC 곡선의 예시를 나타냈다.

4.1.2 학습 방법

Cross-entropy 방법: 합성곱 신경망 기반의 이미지 분류 기술들의 구성을 도식화 하면 **Figure 4.3**과 같다. 각 기술마다 고성능의 이미지 분류를 위한 합성곱 신경망 구조를 제시하고 네트워크 끝단에 완전연결층을 연결함으로써 이미지 분류 데이터셋의 클래스 수와 일치하는 차원의 벡터를 만들어낸다. 이 벡터와 이미지 분류 데이터셋의 실측자료(Ground Truth)를 이용해 cross-entropy 손실함수를 계산하여 역전파를 통해 그 값을 줄이는 방향으로 네트워크를 학습시킨다. cross-entropy 손실함수를 계산하는 방법은 **2.2.1 섹션**에서 소개하였다.

이러한 이미지 분류 기술의 컨셉을 그대로 가져와 도메인을 사람 재인식으로 한정시키고 사람 재인식 데이터셋을 활용함으로써 사람 재인식 네트워크를 학습시킬 수 있다. 이 때, 이미지 분류 데이터셋의 분류하고자 하는 클래스의 개수 C 는 사람 재인식 학습 데이터의 신원 수가 되고, 실측 자료 $\{\mathbf{y}_i\}_{i=1}^N$ 는 사람 재인식 데이터셋의 신원들의 정보를 포함하게 된다.

Triplet 방법: triplet 손실함수 기반의 사람 재인식 방법은 **2.2.2 섹션**에서 기술하였다. 기준, 동일한 사람, 다른 사람 이미지의 쌍을 사람 재인식 학습 데이터셋 τ 로부터 뽑는다. 즉, 모든 $(\mathbf{x}^a, \mathbf{x}^p, \mathbf{x}^n) \in \tau$ 에 대하여, 각각의 데이터 쌍을 사람 재인식 네트워크에 통과시키고 결과로 얻어진 특징 벡터들로 triplet 손실함수를 계산하여 전체 네트워크를 학습시킨다. 여기서 기준, 동일한 사람, 다른 사람의 이미지를 샘플링하는 방법은 [10]을 기반으로 하였다.

4.1.3 테스트 방법

사람 재인식 방법의 테스트 과정은 Cross-entropy 방법, Triplet 방법 모두 동일하다. 완전연결층을 제외한 사람 재인식 네트워크를 통해 사람 재인식 테스트 데이터셋에 포함돼 있는 query 이미지와 gallery 이미지의 특징 지도(feature map)를 뽑고 **Table**

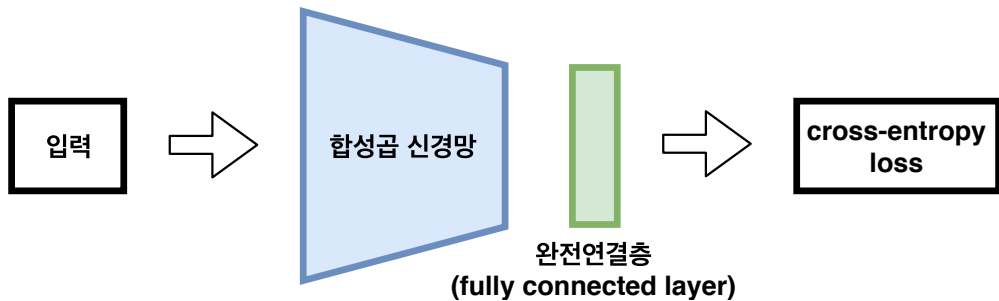


Figure 4.3: 이미지 분류를 위한 딥러닝 네트워크 구조

4.1과 같은 query-gallery 거리 행렬을 만든다. 각 행별로 오름차순 정렬을하면 query 이미지와 거리가 가까운 순으로 gallery 이미지가 정렬이되고, 마지막으로 이 행렬로부터 4.1.1 섹션의 mAP, rank-1, rank-5, rank-10을 계산해서 사람 재인식 알고리즘의 성능을 평가한다.

4.2 실험 결과

4.2.1 실험 세부사항

본 논문에서의 실험은 Cross-entropy 방법, Triplet 방법, 통합 네트워크의 동등한 비교를 위해 추가적인 확장을 붙이지 않고 Chapter3에서 제시한 네트워크 구조만을 구현하여 baseline으로 삼았다. 또한, 사람 재인식 방법론에 관계 없이 대부분의 실험조건을 동일하게 유지하였다. 그 세부 사항으로 데이터셋은 현재 사람 재인식 연구에서 가장 많이 활용되는 Market-1501 [2] 데이터셋을 사용하였고, 네트워크 구현을 위해 Pytorch 라이브러리 [28]를 사용해 실험하였다. 임의 제거는 공통적으로 파라미터를 $p = 0.5$, $s_l = 0.02$, $s_h = 0.4$, $r_1 = 0.3$, $r_2 = 1.0$ 으로 설정하였다. 또한, 손실함수를 통한 그래디언트하강을 수행할 때에는 확률적 그래디언트하강을 사용했다. 학습과정에서 입력 데이터의 batch의 크기는 32로 하였고, 288×144 의 크기로 resize하였으며, 0.5의 확률로 좌우 반전(horizontal flip)되도록 하였다. 마지막으로 triplet 손실함수 기반의 사람 재인식 방법에만 사용되는 margin(α)는 0.3으로 하였다. 마지막으로 모든 실험에서 성능 측정을 위한 메트릭은 4.1.1 섹션의 rank-1(CMC),

rank-5(CMC), rank-10(CMC), mAP을 기반으로 하였다.

4.2.2 Cross-entropy 방법의 비교 실험

기학습 네트워크의 필요성: 2.1.1 섹션에서 소개한 바와 같이 사람 재인식 기술에 적용되는 데이터셋의 크기는 한정적이기 때문에 사람 재인식 네트워크를 구성하는데 있어, 이미지 분류를 목적으로 구현된 합성곱 신경망을 기반으로 세밀 조정하는 방법을 사용한다. 사람 재인식 기술을 구현할 때 이미지 분류 기술을 기반으로 하는 것이 어떤 효과를 보이는지 확인하기 위하여 이미지 분류 데이터셋에 대한 네트워크 기학습 유무에 따른 사람 재인식 기술의 성능 차이를 확인하였다.

Table 4.2: 이미지 분류에 대한 기학습 유무에 따른 사람 재인식 알고리즘의 Market-1501에서 성능 비교

이미지 분류 기반의 사람 재인식 방법				
	rank-1	rank-5	rank-10	mAP
Not pretrained	73.66	89.07	93.02	46.22
Pretrained	86.19	94.92	96.97	67.76

Table 4.2에서 "Not pretrained"의 경우엔 네트워크 가중치의 임의 초기화(random initialization)를 통해 사람 재인식 네트워크를 구성한 것이고, "Pretrained"의 경우엔 이미지 분류 데이터셋에 대해 기학습되지 않은 합성곱 신경망을 이용해 Table 4.2에서 확인할 수 있는 바와 같이 기학습된 이미지 분류 네트워크를 이용하여 내재된 특징점을 활용할 경우 성능향상을 보임을 확인할 수 있다. 사람 재인식에 주로 사용되는 성능 측정 메트릭에 대해 큰 폭의 성능향상을 이루었다. 정리하면 이미지 분류 네트워크의 특징점(feature)이 사람 재인식에 활용될 경우 긍정적인 방향으로 도움을 줄 수 있음을 확인할 수 있다. 본 논문에서 제시한 모든 사람 재인식 네트워크는 이미지 분류 네트워크를 초기상태로 학습되었다.

Cross-entropy 방법에서 롤백의 효과: Cross-entropy 방법에 기반한 사람 재인식 네트워크에 3.4 섹션에서 소개한 롤백을 적용할 경우 어떠한 효과를 보이는지 확인하였다. 이 실험에서는 3.4 섹션의 M 을 4로 두고 학습하였다. 즉, ResNet-50 [16]의 총 4개의 블록 중 첫 번째 블록이 1번, 두 번째 블록이 2번, 세 번째 블록이 3번, 네 번째 블록이 4번 롤백된다.

Table 4.3: Cross-entropy 방법으로 구현된 사람 재인식 네트워크의 성능과 이 네트워크에 롤백을 적용했을 때 Market-1501에서 성능 비교

Cross-entropy 방법으로 구현된 사람 재인식 네트워크				
	rank-1	rank-5	rank-10	mAP
w/o Rollback	86.19	94.92	96.97	67.76
Rollback	90.05	96.44	97.71	73.93

Table 4.3에서 확인 할 수 있듯이 롤백을 적용할 경우 사람 재인식을 위한 합성곱 신경망의 low-level layer의 학습이 강화되어 여러가지 성능 평가 메트릭에 대해 성능 향상을 이루었음을 확인할 수 있다.

4.2.3 Triplet 방법의 비교 실험

3.2 섹션의 Triplet 방법으로 구현된 사람 재인식 기술의 성능과 해당 구조에 3.4 섹션의 롤백을 적용했을 때 성능을 비교하면 Table 4.4와 같다. 롤백은 4.2.2 섹션에서의 방법과 동일하게 수행되었다.

Table 4.4: 기본적인 triplet 손실함수 기반의 사람 재인식 네트워크의 성능과 이 네트워크에 롤백을 적용했을 때 Market-1501에서 성능 비교

Triplet 방법으로 구현된 사람 재인식 네트워크				
	rank-1	rank-5	rank-10	mAP
w/o Rollback	86.10	93.74	96.17	69.66
Rollback	89.16	95.72	97.39	75.37

먼저, 서론에서 소개한 바와 같이 Cross-entropy 방법과 Triplet 방법의 사람 재인식 기술은 나름의 장단이 있으며, Table 4.3와 Table 4.4의 첫 번째 행에서 확인할 수 있는 바와 같이 성능 또한 비슷한 수준으로 나타난다. 추가로, Cross-entropy 방법으로 구현된 사람 재인식 네트워크에서와 마찬가지로 Triplet 방법으로 구현된 사람 재인식 네트워크에도 롤백을 적용할 경우 성능향상을 보임을 알 수 있다. 모든 성능 측정 메트릭에 대해 성능 향상을 보였다. Table 4.4의 결과는 Triplet 방법으로 구현된 사람 재인식 네트워크에서 추가적인 확장(add-on)없이 단순한 학습 방법의 개선만으로 성능 향상을 이루어낼 수 있음을 보였고, 네트워크에 추가적인 확장이 된다면 더 높은 성능을 얻을 수 있는 가능성을 시사한다.

4.2.4 통합 네트워크의 비교 실험

사람 재인식의 두가지 방향성인 Cross-entropy 방법과 Triplet 방법을 통합하여 통합 네트워크(Integrated Network)를 구성할 수 있다. 이는 3.3 섹션에서 소개하였다. Figure 3.3의 구조를 가지는 통합 네트워크를 통해 사람 재인식을 수행할 경우의 성능과 Cross-entropy 방법과 Triplet 방법을 각각 따로 수행한 경우의 성능을 비교하면 Table 4.5와 같다.

Table 4.5: 통합 네트워크의 Market-1501에서 성능 비교

통합 네트워크로 구현된 사람 재인식 네트워크				
	rank-1	rank-5	rank-10	mAP
Cross-entropy 방법	86.19	94.92	96.97	67.76
Triplet 방법	86.10	93.74	96.17	69.66
통합 네트워크	87.52	95.72	97.47	71.33

Table 4.5의 결과를 보면 Cross-entropy 방법, Triplet 방법의 사람 재인식 방법을 각각 따로 수행할 때 보다 두 가지를 방향성을 통합하여 하나의 네트워크로 구성할 경우 성능향상을 보임을 확인할 수 있다.

4.2.5 통합 네트워크에서 롤백의 효과

사람 재인식 네트워크에서 롤백을 적용하면 성능 향상을 보임을 4.2.2 섹션과 4.2.3 섹션에서 확인하였다. 이를 통해 사람 재인식 방법의 두 가지 방향성을 하나로 통합한 네트워크에 롤백을 적용할 경우 성능 향상을 이룰 것임을 예측할 수 있고, 성능 변화는 다음과 같다.

Table 4.6: 통합 네트워크의 성능과 이 네트워크에 롤백을 적용했을 때 Market-1501에서 성능 비교

통합 네트워크로 구현된 사람 재인식 네트워크				
	rank-1	rank-5	rank-10	mAP
w/o Rollback	87.52	95.72	97.48	71.33
Rollback	90.11	96.08	97.60	76.50

통합 네트워크에 롤백을 적용할 경우 Cross-entropy 방법, Triplet 방법에서 처럼

모든 메트릭에 대해 성능 향상을 보인다. 즉, 사람 재인식의 두 가지 방향성은 하나로 통합이 가능하며 그 성능 또한 각각을 따로 수행할 때 보다 높음을 확인할 수 있다.

4.2.6 실험 결과 종합

본 논문에서는 사람 재인식 기술의 두 가지 방향성인 Cross-entropy 방법과 Triplet 방법을 통합하였고 이러한 통합 네트워크에 대해 롤백의 효과를 검토하였다. 4.2 섹션에서 수행된 모든 실험 결과들을 정리하면 다음과 같다.

Table 4.7: 본 논문에서 제시한 네트워크들의 Market-1501에서 성능 비교

	rank-1	rank-5	rank-10	mAP
Not pretrained	73.66	89.07	93.02	46.22
Cross-entropy based Re-ID(w/o Rollback)	86.19	94.92	96.97	67.76
Cross-entropy based Re-ID with Rollback	90.05	96.44	97.71	73.93
Triplet based Re-ID(w/o Rollback)	86.10	93.74	96.17	69.66
Triplet based Re-ID with Rollback	89.16	95.72	97.39	75.37
Integrated Network(w/o Rollback)	87.52	95.72	97.48	71.33
Integrated Network with Rollback	90.11	96.08	97.60	76.50

Table 4.9의 2, 4, 6번째 행으로부터 Cross-entropy 방법, Triplet 방법을 통합할 경우 높은 성능을 얻어 낼 수 있음을 확인할 수 있다. 추가로, 이러한 통합 네트워크에 롤백을 적용할 경우 mAP와 rank-1 메트릭에 대해 가장 높은 성능을 달성하였는데, 이는 사람 재인식 분야의 성능 메트릭 중 가장 자주 언급되는 메트릭이다. 이를 통해 통합 네트워크에 롤백을 적용할 경우 가장 높은 성능을 달성할 수 있음을 확인할 수 있다.

4.2.7 각 scheme 별 통합네트워크의 성능

[15]에서는 Cross-entropy 방법의 사람 재인식 기술에 롤백이 적용될 때, 각각의 scheme별로 점증적인 성능 향상이 이루어짐을 보였다. 해당 섹션에서는 통합 네트워크에 롤백이 적용될 때, 각 scheme별로 어떠한 성능을 보이는지 확인하였다. 이 실험에서 scheme1은 3.4 섹션의 M 을 1로, scheme2는 M 을 2로, scheme3은 M 을 3으로, scheme4는 M 을 4로 설정한 것이다.

Table 4.8: 통합 네트워크의 scheme별 성능

통합 네트워크로 구현된 사람 재인식 네트워크				
	rank-1	rank-5	rank-10	mAP
scheme1	87.52	95.72	97.48	71.33
scheme2	89.01	95.98	97.54	73.86
scheme3	89.86	96.06	97.58	75.46
scheme4	90.11	96.08	97.60	76.50

Table 4.8로부터 통합 네트워크에서도 롤백 scheme별로 점증적인 성능 향상을 보임을 확인할 수 있고, 4번째 scheme에서 가장 높은 성능을 보임을 확인할 수 있다.

4.2.8 기존에 가장 높은 성능을 발휘하는 사람 재인식 네트워크들과 통합 네트워크의 성능 비교

기존에 사람 재인식 분야에는 다양한 연구가 존재한다. 해당 섹션에서는 기존에 가장 높은 성능을 발휘했던 사람 재인식 기술들과 본 논문에서 제시한 통합 네트워크의 성능을 비교하였다. 그 결과는 **Table 4.9**와 같다.

Table 4.9: 기존 최고 성능(state-of-the-art)의 사람 재인식 기술과 통합 네트워크의 Market-1501에서 성능 비교

	rank-1	mAP
SVDNet [29]	82.3	62.1
PDC [30]	84.1	63.4
PT-GAN [31]	79.8	58.0
AACN [32]	85.9	66.9
HAP2S_P [33]	84.6	69.4
PSE [34]	87.7	69.0
CamStyle [35]	89.2	71.6
PN-GAN [36]	89.4	72.6
MGCAM [37]	83.8	74.3
Ours(Integrated Network)	90.11	76.50

Table 4.9에서는 사람 재인식에서 가장 빈번하게 사용되는 성능 측정 메트릭인 rank-1과 mAP로 그 성능을 비교하였다. 두 성능 측정 메트릭 모두 통합 네트워크에서 가장 높은 성능을 달성함을 확인할 수 있다.

Chapter 5

결론

본 논문에서는 사람 재인식의 두 가지 방향성인 Cross-entropy 방법과 Triplet 방법을 소개하였고 이를 비교하고 설명하였다. 각 방법들의 고유한 특성 때문에 어떤 방법이 사람 재인식 문제를 해결하는데 유력한 방법이다라고 말할 수 없다. 따라서, 사람 재인식 문제를 해결할 때는 두 방법을 함께 고려하는 것이 옳다. 하지만 사람 재인식 분야에서 학습 방법 개선을 위한 방법론인 롤백에 관한 기존 연구는 이 중 한가지 방법인 Cross-entropy 방법에 대한 효용성만 입증하였다. 따라서, 또 다른 방법인 Triplet 방법의 사람 재인식에도 롤백을 적용함으로써 성능향상을 이룰 수 있음을 보였다. 추가적으로, 본 논문에서는 두 가지 방법을 통합하여 하나의 네트워크로 구성하는 방안을 제시하였으며 각각을 따로 수행했을 때보다 높은 성능을 발휘하였다. 또한, 이 방법에 대해서도 롤백을 적용하여 성능향상을 이루었다.

Bibliography

- [1] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3754–3762, 2017.
- [2] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1116–1124, 2015.
- [3] Wei Li, Rui Zhao, and Xiaogang Wang. Human reidentification with transferred metric learning. In *Asian conference on computer vision*, pages 31–44. Springer, 2012.
- [4] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. Person re-identification by descriptive and discriminative classification. In *Scandinavian conference on Image analysis*, pages 91–102. Springer, 2011.
- [5] Douglas Gray and Hai Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European conference on computer vision*, pages 262–275. Springer, 2008.
- [6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

- [7] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Deep metric learning for person re-identification. In *2014 22nd International Conference on Pattern Recognition*, pages 34–39. IEEE, 2014.
- [8] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 152–159, 2014.
- [9] Zhedong Zheng, Liang Zheng, Zhilan Hu, and Yi Yang. Open set adversarial examples. *CoRR*, abs/1809.02681, 2018.
- [10] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.
- [11] Chi Su, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Deep attributes driven multi-camera person re-identification. In *European conference on computer vision*, pages 475–491. Springer, 2016.
- [12] Ruimao Zhang, Liang Lin, Rui Zhang, Wangmeng Zuo, and Lei Zhang. Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification. *IEEE Transactions on Image Processing*, 24(12):4766–4779, 2015.
- [13] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016.
- [14] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.
- [15] Youngmin Ro, Jongwon Choi, Dae Ung Jo, Byeongho Heo, Jongin Lim, and Jin Young Choi. Backbone can not be trained at once: Rolling back to pre-trained network for person re-identification. *CoRR*, abs/1901.06140, 2019.

- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [17] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: a deep quadruplet network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2017.
- [18] Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu. Deep mutual learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4320–4328, 2018.
- [19] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. Svdnet for pedestrian retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3800–3808, 2017.
- [20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [21] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [22] Niall McLaughlin, Jesus Martinez Del Rincon, and Paul Miller. Data-augmentation for reducing dataset bias in person re-identification. In *2015 12th IEEE International conference on advanced video and signal based surveillance (AVSS)*, pages 1–6. IEEE, 2015.
- [23] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*, 2017.

- [24] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [25] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camstyle: a novel data augmentation method for person re-identification. *IEEE Transactions on Image Processing*, 28(3):1176–1190, 2019.
- [26] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [27] D Randall Wilson and Tony R Martinez. The general inefficiency of batch training for gradient descent learning. *Neural networks*, 16(10):1429–1451, 2003.
- [28] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.
- [29] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. Svdnet for pedestrian retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3800–3808, 2017.
- [30] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Pose-driven deep convolutional model for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3960–3969, 2017.
- [31] Jinxian Liu, Bingbing Ni, Yichao Yan, Peng Zhou, Shuo Cheng, and Jianguo Hu. Pose transferrable person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4099–4108, 2018.

- [32] Jing Xu, Rui Zhao, Feng Zhu, Huaming Wang, and Wanli Ouyang. Attention-aware compositional network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2119–2128, 2018.
- [33] Rui Yu, Zhiyong Dou, Song Bai, Zhaoxiang Zhang, Yongchao Xu, and Xiang Bai. Hard-aware point-to-set deep metric for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 188–204, 2018.
- [34] M Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 420–429, 2018.
- [35] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camera style adaptation for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5157–5166, 2018.
- [36] Xuelin Qian, Yanwei Fu, Tao Xiang, Wenxuan Wang, Jie Qiu, Yang Wu, Yuguang Jiang, and Xiangyang Xue. Pose-normalized image generation for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 650–667, 2018.
- [37] Chunfeng Song, Yan Huang, Wanli Ouyang, and Liang Wang. Mask-guided contrastive attention model for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1179–1188, 2018.

초 록

사람 재인식 (Person re-identification) 기술은 여러 개의 겹치지 않는 카메라 뷰 (camera view)로부터 촬영된 사람들의 정보 중 동일한 사람의 정보를 매칭시키는 기술이다. 이러한 기술들은 대인 이동 경로 추적, 감시 시스템(visual surveillance), 포렌식(forensic) 등에 유용하게 적용될 수 있다. 사람 재인식 기술은 크게 두가지 방향성을 가지는데, 이미지 분류 기반의 cross-entropy 손실함수를 사용하는 방법과 이미지 쌍을 입력받아 triplet 손실함수를 사용하는 방법이 있다. 각각의 방법들의 고유한 특성 때문에 어떤 방법이 열린 집합 문제를 해결하는데 더 유력한 방법이라고 비교하기는 어렵다. 본 논문에는 이미지 분류 기반의 방법과 triplet 손실함수 기반의 방법을 비교하고, 기존에는 이미지 분류 기반의 방법에만 활용되었던 롤백 (Rollback)의 효과를 triplet 손실함수 기반의 방법에도 적용하여 해당 방법에 대한 롤백의 유용성을 검토하였다. 추가로, 사람 재인식 기술의 두 방법을 동시에 적용하여 더 좋은 성능을 내는 알고리즘을 제안하였다.

주요어: 사람 재인식, 롤백, 트리플렛 손실 함수, 이미지 분류

학번: 2017-20953

감사의 글

대학원 생활을 시작한 지 2년, 저는 여전히 많이 부족하지만 찌는듯한 더위가 두 번이나 찾아온 것을 보니 세월의 흐름을 느끼게 됩니다. 그렇게 길지 않은 시간이었지만 여러가지 우여곡절도 많았던 제 석사과정 중에 주변 사람들의 도움이 없었다면 이렇게 값진 성과를 얻어내지 못했을 것입니다. 그 고마움을 다 표현할 수는 없겠지만 이렇게 감사의 글로나마 그 마음을 남겨봅니다.

먼저 최진영 지도교수님께 감사 인사를 드립니다. 교수님께는 감사한 마음과 동시에 죄송스런 마음도 듭니다. 많이 부족한 지도학생인 제게 항상 관심을 가져주시고 걱정해주시는 덕에 무사히 석사과정을 마치고 졸업을 할 수 있게 된 것 같습니다. 이제 곧 연구생 소속에서 벗어난다는 것에 불안하고 초조하지만 교수님께서 제게 보여주셨던 모습처럼 매사에 최선을 다하도록 하겠습니다.

다음으로 인지능연구실 재학생 여러분께도 감사하다는 말을 전하고 싶습니다. 지금은 졸업하고 안계시지만 많은 조언을 해주셨던 문섭이 형, 상두 형, 종원이 형, 동갑이지만 선배로서 배울 점이 많은 병호, 힘들 때마다 힘이 되는 조언을 해주신 기경이 형, 병주 형, 까마득한 후배인데 나이만 많아서 힘들었을 지엽, 영한, 준호, 연구실 생활에 많은 도움을 주셨던 영민이 형, 같이 수업들으며 대학원 생활에 대한 많은 조언을 해주었던 종인, 대웅, 지웅, 같이 지냈던 기간이 짧아 아쉬운 대홍이 형, 호석이 형, 그리고 함께 입학해 아무것도 모르는 연구실 생활 초기 힘이 됐던 규왕, 현준, 함께 석사과정을 보내 많은 도움이 됐던 형길이 형, 태길, 선오, 동기도 없이 힘들지만 연구실에 잘 적응해나가는 모습에 배울점이 많은 대호, 졸업하는데 도움만 받아 미안한 슬기누나 모두에게 감사드립니다. 연구실 생활 내내 함께해 많은 추억을 쌓았는데, 떨어지게 된다니 벌써부터 아쉬움이 많이 남습니다. 앞으로의 인생에 지금처럼 좋은 사람들을 또 만날 수 있을까하는 걱정이 듭니다. 제 연구실 생활에 큰

힘이 됐던 인지능연구실 재학생 분들께 다시 한 번 감사의 말씀을 전합니다.

마지막으로 사랑하는 가족들에게도 감사의 말씀을 전해드립니다. 대학원 생활 뿐 아니라 제 일생동안 든든한 버팀목이 되어주신 아버지, 어머니, 늦은 나이까지 공부하는 형을 위해 많은 양보를 해준 동생 상혁, 모두에게 감사드립니다. 가족들에게는 항상 감사하지만 표현을 하지 못했는데, 이 감사의 글을 빌어 다시 한번 감사하다는 말씀을 드립니다.