깊은 인공 신경망 이미지 기술자를 활용하는 멤버 분류

장영균, 이석희, 조남익
서울대학교 전기정보공학과
{kyun0914, seokheel}@ispl.snu.ac.kr, nicho@snu.ac.kr

Member Verification with Deep Learning-based Image Descriptors

Young Kyun Jang, Seok Hee Lee and Nam Ik Cho

요 약

최근 딥 러닝을 이용한 방법들이 이미지 분류에서 뛰어난 성능을 보임에 따라, 복잡한 특징을 담고 있는 얼굴 이미지에 대해 이를 적용하려는 시도가 늘어나고 있다. 특히, 이미지로부터 주요한 특징들을 추출하여 간결하게 이미지를 대표할 수 있는 이미지 기술자 (Image descriptor)를 딥 러닝을 통해 생성하는 연구가 인기를 끌고 있다. 이는 딥 러닝 끝 단에 있는 Fully-connected layer 의 출력으로 얻을 수 있으며 이미지의 의미론적 상관관계를 이용하여 학습된다. 구체적으로, 이미지 기술자는 실수형 벡터 데이터로서, 한 장의 이미지를 수치화 하여 비슷한 이미지 사이에는 벡터 거리가 가깝게, 서로 다른 이미지 사이에는 벡터 거리가 멀게 구성된다. 본 연구에서는 미리 학습된 인공 신경망을 통과시켜 얻은 얼굴 이미지 기술자를 활용하여 멤버 분류를 위한 두 개의 인공 신경망을 학습하는 것을 목표로 한다. 제안된 방법을 검증하기 위해 얼굴 인식에 널리 사용되는 벤치 마크 데이터셋을 활용하였고, 그 결과 제안된 방법이 높은 정확도로 멤버를 분류할 수 있다는 것을 확인하였다.

1. 서론

멀티미디어와 카메라 기술이 발전함에 따라, 매일 새로운 이미지가 기하 급수적으로 생성되고 있다. 이에 따라, 이를 활용할 수 있는 여러 딥 러닝 기반의 컴퓨터 비전 작업들이 주목받고 있는데, 그 중 가장 중요한 것으로는 이미지를 분류하는 작업이 있다. 본 논문에서 다루는 이미지 분류는, 이미지가 어떤 사람인지, 혹은 어느 그룹에 속하는지를 나누는 작업에 중점을 두며, 이는 실 생활에서 유용하게 사용될 수 있는 중요한 분야이다. 본 연구에서는, 이미지를 그룹의 멤버로 등록하고, 쿼리 이미지가 그 그룹에 속하는 지를 분류할 수 있는 멤버 분류기를 학습하는 것을 목표로 한다.

이미지 기술자 (Image descriptor)란, 이미지를 대표할 수 있는 짧은 코드로서, 대표적으로 SIFT (Scale Invariant Feature Transfrom) [1]나 SURF (Speed-Up Robust Features) [2]와 같은 알고리즘을 적용하여 얻을 수 있다. 그러나 이러한 이미지 기술자들은 RGB 각 채널의 고유한 특징이 아닌

Grayscale 이미지를 활용한 것이므로, 분류 성능에 한계를 가진다. 게다가, 이들은 이미지의 레이블 정보를 활용하지 않고 있어서 정확한 분류 성능을 기대하기 어렵다는 단점 또한 가지고 있다. 반면, 이미지의 레이블 정보를 최대한 활용하는 딥 러닝 기반의 이미지 기술자들은 이미지 분류 영역에 있어서 기존의 방식들 대비 훨씬 뛰어난 성능을 보여주고 있으므로, 본 연구에서도 이를 활용하여 성능이 높은 멤버 분류기를 만들고자 한다.

2. 데이터 셋

그림 1. LFW 데이터 셋 예시

본 연구에서 활용하고자 하는 데이터 셋은 얼굴 이미지 분류에서 널리 쓰이는 Labeled Faces in the Wild (LFW) [3] 데이터 셋으로, 그림 1 은 데이터 셋에 담긴 예시 이미지들을 나타낸다. 본 데이터 셋은 총 13,233 장의 이미지로 이루어져 있으며 이는 5,749 명의 사람의 이미지를 담고 있다. 그리고 예시 이미지와 같이, 성별, 나이, 인종의 구별 없이 다양한 사람의 얼굴 이미지들을 담고 있다. 그러나 예시 이미지를 살펴보면, 한 장의 이미지에 다른 사람의 얼굴이 부분적으로 나타나는 경우도 있고, 얼굴 이외의 다른 물체나 배경이 존재하는 경우도 있으므로 이를 직접적으로 멤버 분류기를 학습하기 위해 사용하기엔 어려움이 있다.

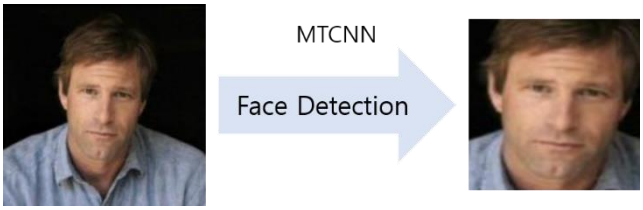

그림 2. 얼굴 영역 검출 및 정렬 예시

따라서 본 연구에서는 정확한 얼굴 이미지를 멤버 분류기 학습에 사용하기 위해, 그림 2 와 같이 딥 러닝 기반의 얼굴 추출기(Face Detector)인 MTCNN [4] 알고리즘을 LFW 데이터 셋에 적용하여, 얼굴 영역을 찾아내어 이를 활용한다. 이 과정에서, 더 높은 정확도를 얻기 위해, 찾아낸 얼굴 영역에 대해 이를 정면으로 정렬하는 알고리즘도 포함되어 적용된다.

3. 이미지 특징 추출 및 멤버 분류기 학습

인공 신경망을 활용하여 이미지의 특징을 추출하기 위한 연구도 활발히 이루어 지고 있다. 그 중 가장 대표적인 방식은 인공 신경망의 Fully connected layer 의 출력인 피쳐 벡터를 이미지 기술자로 활용하는 방법이다. 이렇게 얻은 이미지 기술자는, 인공 신경망의 심층 구조에서 추출된 다양한 이미지 피쳐를 압축하여 실수 값의 벡터 형태로 표현하고 있다. 이미지 레이블을 사용하여 이미지 사이의 의미론적 상관관계를 찾고, 이를 활용하는 적절한 목적함수를 통해 인공 신경망을 학습한다면 이미지 기술자 사이에서 높은 분류 성능을 기대할 수 있다.

인공 신경망 피쳐 추출기로는 FaceNet [5] 알고리즘을 통해 학습된 모델을 사용한다. FaceNet 알고리즘은 이미지로부터 고정된 크기의 이미지 기술자를 인공 신경망을 통해 추출한 뒤, 이를 삼중항 목적함수 (Triplet Loss)를 통해 학습하여, 의미론적 상관관계를 이미지 기술자에 담아내는 것을 목표로 학습한다. 이에 사용되는 목적 함수는 아래의 식 (1)과 같다.

$$\boldsymbol{L_{Triplet}} = \mathbf{d}(\mathbf{a}, \mathbf{p}) - \mathbf{d}(\mathbf{a}, \mathbf{n}) + \boldsymbol{m} \qquad (1)$$

이 때, d(·,·)는 유클리디안 거리, **a**는 앵커 이미지 기술자, **p**는 앵커와 같은 레이블을 갖는 샘플의 이미지 기술자, **n**은 앵커와 다른 레이블을 갖는 샘플의 이미지 기술자를 의미하고, m은 이 둘의 거리 사이에 여유 공간을 주기 위한 양의 실수 값이다. 이 때, 피쳐 추출기의 구조로는 Inception-ResNet v1 [6]가 사용되었으며, 데이터 셋으로는 약 300 만 장의 이미지로 구성된 VGG-FACE2 [7]가 사용되었다.

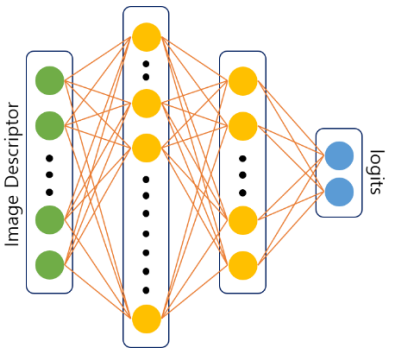

그림 3. 멤버 분류기 예시

인공신경망 기반 이미지 기술자는 위의 특징 추출기를 활용하여 얻어지는데, 이는 128 차원의 실수 값 벡터의 형태를 띄고 있다. 본 연구에서는 이를 활용하여 원하는 사람들끼리 그룹을 짓고, 쿼리 이미지가 그 그룹에 속하는지, 속하지 않는 지를 구분할 수 있는 멤버 분류기를 학습한다. 멤버 분류기의 입력이 벡터이기 때문에, 이를 활용하기 위해서 그림 3 과 같이 Fully connected layer 세 개와 2 개의 ReLU (Rectified Linear Unit) 활성화 함수를 사용하여 멤버 분류기를 구성한다. 이 때, 각 Fully connected layer 의 차원은 [128×256], [256×128], [128×2] 로 출력으로 멤버인지 아닌지의 확률 값인 logits 를 생성한다. 여기에 추가적으로 Softmax 함수를 적용하여 logits 값을 0~1 사이의 확률 값($\widehat{y_{ik}}$)으로 매핑하는 과정을 거친다. 얻어낸 확률 값과 멤버 레이블을 활용하여 아래의 식 (2)의 Binary Softmax Cross entropy(BCE loss)를 계산하고, 이를 최소화하는 방향으로 멤버 분류기를 학습한다.

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{2} y_{ik} log\ \widehat{y_{ik}} \qquad (2)$$

4. 실험 결과 및 분석

실험은 두 개의 멤버 분류기를 학습하는 것을 목표로 구성된다. 본 연구에서는 LFW 데이터 셋의 13,233 장의 이미지를 두 개로 그룹 지어서, A 그룹은 2,749 명의 5,000 장의 이미지로 구성하고, B 그룹은 남은 3,000 명의 8,000 이미지로 구성하였다.

A 그룹 멤버 분류기는 5,000 장의 이미지를 멤버로 구분하고, B 그룹의 이미지와 VGG-FACE2 데이터 셋에서 랜덤하게 추출한 5,000 장의 이미지를 멤버가 아닌 이미지로 학습한다. 마찬가지로 B 그룹 멤버 분류기는 8,000 장의 이미지를 멤버로 구분하고, A 그룹의 이미지와 앞과 동일한 VGG-FACE2 의 이미지로 학습한다. 성능 평가는 정확도 (Accuracy, %)로 측정하며, 각 멤버 분류기가 학습에 사용한 데이터 셋 모두를 인증 시도로 사용하여(각 5,000 번, 8,000 번) 정확도를 측정한다.

제안하는 방법과의 비교를 위해, 비지도학습 기반의 클러스터링 방식인 K-means clustering 과 다른 지도 학습 기반인 LIBSVM [8]을 활용한 분류기를 구성하여 멤버 분류 실험을 진행하였다. 이 둘은 우리가 제안하는 방법과 동일한 이미지 기술자를 사용하여 멤버 분류기를 학습한다.

	Group A	Group B
Ours	92.2%	91.6%
K-means	58.8%	52.8%
LIBSVM [8]	84.2%	83.0%

표 1. 신경망 성능 실험 결과

표 1 에 따르면, 제안하는 방법이 다른 알고리즘들에 비해 월등한 성능으로 멤버 분류를 수행하는 것을

확인 할 수 있다. K-means 방식과의 차이는 지도학습과 비지도학습의 차이에 의하여 발생하는 것을 확인할 수 있다. 그리고 LIBSVM 과의 차이는 인공 신경망이 멤버 분류기를 학습하는데 더 최적화되어 있다는 것을 의미한다.

5. 결론

본 연구에서는 인공 신경망 기반의 이미지 기술자를 효과적으로 그룹화 하여 멤버 분류에 사용할 수 있는 방법을 제안하였다. 이 방법을 활용하면, 인공 신경망 기반 이미지 기술자를 쉽게 원하는 대로 그룹지을 수 있다는 장점을 가진다.

감사의 글

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 1711075689, AI 어플리케이션을 지원하는 IoT 연동 분산 Edge 클라우드 기술 개발)

참고문헌

[1] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." , IEEE Transactions on pattern analysis and machine intelligence 24.7, 2002

[2] H. Bay et al., “Speeded-up robust features (SURF),” Computer Vision and Image Understanding, 2008.

[3] Huang, Gary B., and Erik Learned-Miller. "Labeled faces in the wild: Updates and new reporting procedures." Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep (2014): 14-003.

[4] Zhang, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." IEEE Signal Processing Letters 23.10 (2016): 1499-1503.

[5] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[6] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Thirty-first AAAI conference on artificial intelligence. 2017.

[7] Cao, Qiong, et al. "Vggface2: A dataset for recognising faces across pose and age." 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, 2018.

[8] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A library for support vector machines." ACM transactions on intelligent systems and technology (TIST) 2.3 (2011): 1-27.