

개인정보 보호를 위한 비디오에서의 지능형 얼굴 모자이크 방법

임혁, 최민석, 최승비, 최해철
한밭대학교 정보통신공학과

skdltm117@nate.com, smope96@naver.com, sseung_bi@naver.com, choihc@hanbat.ac.kr

Intelligent Face Mosaicing Method in Video for Personal Information Protection

Hyuk Lim, Minseok Choi, Seungbi Choi, and Haechul Choi
Dept. of Information Community Engineering, Hanbat National University

요약

개인 방송의 보편화로 인해 인터넷 혹은 방송으로 유포되는 영상에서 일반인의 얼굴이 빈번히 노출되고 있으며, 동의 받지 않은 얼굴의 방송 노출은 개인 초상권 침해와 같은 사회적 문제를 일으킬 수 있다. 이러한 개인 초상권 침해 문제를 해결하고자 본 논문은 비디오에서 일반인의 얼굴을 검출하고 이에 마스크를 가하는 방법을 제안한다. 제안 방법은 우선 딥러닝 기반의 Faster R-CNN을 이용하여 모자이크를 하지 않을 특징인과 모자이크를 가할 비특정인을 포함한 다수의 얼굴 영상을 학습한다. 학습된 네트워크를 이용하여 입력 비디오에 대해 사람의 얼굴을 검출하고 검출된 결과 중 특정인을 선별해 낸다. 최종적으로 입력 비디오에서 특정인을 제외한 나머지 검출된 얼굴에 대해 모자이크 처리를 수행함으로써 비디오에서 지능적으로 비특정인의 얼굴을 가린다. 실험결과, 특정인과 비특정인을 포함한 얼굴 검출의 경우 99%의 정확도를 보였으며, 얼굴 검출 결과 중 특정인을 정확히 맞춘 경우는 86%의 정확도를 보였다. 제안 방법은 인터넷 동영상 서비스 및 방송 분야에서 개인 정보 보호를 위해 효과적으로 활용될 수 있을 것으로 기대된다.

I. 서론

최근들어 OTT(Over The Top) 서비스라 불리는 인터넷 기반 동영상 서비스가 확장되고 업로드되는 영상의 수가 증가함에 따라 개인정보 보호법을 위반한 영상의 빈도가 증가했다. 이는 개인이 영상을 편집할 때 영상에 나오는 일반인에 대한 완벽한 모자이크 처리가 어렵기 때문에 Deep Learning을 통해 특정인을 제외한 나머지 인물들을 자동으로 모자이크 처리하는 서비스를 구현하여 이와 같은 문제를 해결하고자 한다.

단일 영상을 입력 값으로 넣어주어 학습을 시키고 이를 통하여 생성된 모델을 통해 특정인과 비특정인을 식별하여 자동으로 모자이크 처리해주는 프로그램을 구현할 것이다. Faster-RCNN을 통해 식별하고자 하는 객체의 Feature Map을 추출하여 RPN(Region Proposal Network)를 생성하는데 이 과정에서 Representational Bottleneck 현상이 발생할 수 있다. 이를 해결하기 위해 InceptionV2를 사용해 Feature Map을 효과적으로 줄일 수 있도록 Convolution과 Pooling연산을 병렬적으로 수행해 RPN을 생성하고 Proposal을 제안한다. Proposal과 사진데이터에서 추출한 GT(Ground Truth)를 비교하여 IoU(Intersection over Union)를 계산해 설정한 임계값 이상이면 객체탐지에 성공한 것으로 판단해 특정인과 특정인과 비특정인을 구별해낸다. 구별된 비특정인에는 자동적으로 모자이크 처리를 함으로써 개인정보를 보호할 수 있다.

II. 본론

2.1 Data Set

연구의 목적을 달성하기 위해 인종, 연령을 제한하지 않고 다양한 인물의 데이터를 크롤링해 약 24,000장의 데이터를 확보하였고, 특정인 3명의 데이터는 직접 촬영해 2,500장의 데이터를 확보하였다. 이 중 21,432장을 training data로 5,360장을 test data로 사용하였다.

정확한 Object detection을 위해 데이터 셋을 만들 때 각각의 특정인과 비특정인을 구분하기 위한 라벨을 달아주어야 한다. 비특정인은 person으로, 3명의 특정인은 각각의 이름에서 이니셜을 사용하여 라벨을 달아 Labelmap에 등록시켜 Training과 Test를 진행할 때 인물 영상에서 특정인을 분류할 수 있도록 구현하였다. 이를 통해 하나의 프레임에 특정인과 비특정인이 같이 존재한다면 Labelmap을 통하여 각각을 분류하여 검출할 수 있도록 하였다.

2.2 Faster-RCNN

Faster-RCNN은 CNN(Convolution Neural Network)을 기반으로 개선되고 발전된 Object Detection 네트워크 중 하나이다. 2014년 발표된 RCNN의 속도측 문제점을 개선하여 2015년에 Fast-RCNN이 발표되었고 모델에서 Region Proposal 단계의 Selective Search로 하던 후보영역 생성을 RPN(Region Proposal Network)으로 대신하여 속도를 한층 더 개선시킨 Faster-RCNN이 발표되었다. Faster-RCNN은 다른 알고리즘들보다 좋은 식별 능력을 보여준다. 이 논문에서 특정인과 비특정인을 정확하게 식별하여 모자이크 처리하는 것이 가장 중요하다 생각하여 Faster-RCNN을 사용하였다.

Faster-RCNN은 입력 영상에 CNN을 통해 특징점을 추출하여 feature map을 만들고 최종 feature map을 RPN의 입력으로 받아들인다. 이후 $n \times n$ 크기의 sliding window convolution을 수행하여 256차원 또는 512차원의 벡터를 만들어내고 이 벡터를 검출하고자 하는 Object인지 아닌지를 나타내는 Box Classification Layer와 후보 영역의 좌표를 만들어내는 Box Regressor Layer에 연결한다. 앞의 두 계층에서 얻은 값들로 RoI(Region of Interests)를 계산 후 다시 처음의 feature map에 오버랩해 RoI Pooling을 적용하고 Classification에 적용하면 Object의 종류를 알아낼 수 있게 된다.

아래 그림은 Faster-RCNN의 구조를 나타낸다.

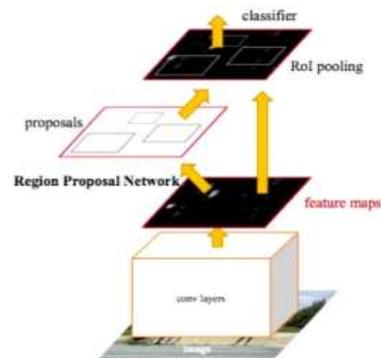


그림 1. Faster-RCNN의 구조 [1]

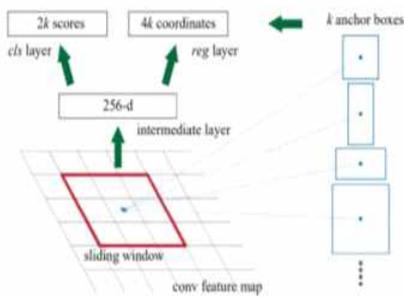


그림 2. Region Proposal Network의 구조 [1]

2.3 InceptionV2 [2]

구글에서 발표한 22개의 Layer로 구성된 CNN 구조의 심층 신경망인 GoogLeNet 모형의 가장 중요한 기술 중 하나인 Inception 모듈은 신경망을 더 넓고 깊게 구성함에도 불구하고 모델의 연산량이 추가적으로 늘지 않고 정확도는 향상한다는 장점이 있다. 현재 Inception 모듈은 V4까지 개발되었지만 이 논문에서는 안정적이면서 많은 데이터를 얻을 수 있는 Inception V2 모듈을 사용하였다. Inception V2는 2015년 구조 변화를 통해 3개의 Convolution으로 구현되고, stride를 2로 적용하여 Pooling 효과를 얻도록 수정되어 총 42개의 Layer로 깊어지게 되었지만 연산량은 2.5배 늘어난 수준으로 여전히 효율성을 가지고 있다.

다수의 필터를 통해 입력 값에 대해 Convolution과 Pooling에 대한 연산을 병렬적으로 수행함으로써 Representational Bottleneck 현상을 해결하였고, VGGNet를 참고하여 5x5 Convolution 연산 1번을 3x3 Convolution 연산 2번으로 변경함으로써 동일한 효과에 대해 연산량을 절약하여 Factorization 문제를 해결하였다.

III. 결과

학습된 네트워크 모델의 성능을 파악하기 위해 준비한 Data Set의 20%를 실험데이터로 사용하여 Test를 진행하였다. 이 실험데이터로 모델을 평가하기 위한 성능은 IoU(Intersection of Union) 값을 사용하였다. IoU는 영상에서 결과를 직접 도출해내 만든 GT의 Bounding Box와 네트워크 모델이 예측한 Bounding Box를 비교했을 때 얼마나 많이 겹치는가로 평가하게 된다. 많이 겹칠수록 정답에 가까운 것이라고 판단되며 일반적으로 IoU Threshold를 0.5로 설정한다.

본 논문에서는 더 정확한 결과를 얻고자 IoU Threshold를 0.8 즉, Bounding Box가 80%이상 겹쳤을 때 예측이 성공했다고 판단하도록 설정했으며, 검출하고자 하는 객체의 Class 각각의 Training Data의 양이 너무 적거나 많아도 정확도가 떨어지는 것을 확인했다. 비특정인의 경우 99%라는 상당히 좋은 결과를 얻었지만 특정인의 경우 평균 86%의 정확도를 얻었다. 검출된 인물의 Label을 구분하여 조건문을 통해 비특정인에게 모자이크 처리를 해줌으로써 개인정보 보호를 위한 프로그램을 구현시켰다.



그림 3. 결과 화면 예시 (비특정인 모자이크)

IV. 결론

본 논문에서는 영상에서 특정인과 비특정인을 검출하여 모자이크 처리하는 프로그램을 제안하였다. 이를 위해서 사용한 신경회로망 네트워크인 Faster-RCNN은 YOLO, SSD와 같은 다른 네트워크에 비해 속도가 느리지만, 정확도가 높기 때문에 OTT서비스에 적합하다고 판단하여 사용하였다. 하지만, 본 논문에서 연구한 결과에 따르면 특정인을 검출한 결과의 정확도가 비교적 낮으므로 충분한 양의 Data를 학습시키고 학습된 모델의 손실률을 낮추는 등 프로그램을 보완한다면 더 정확한 결과를 낼 수 있게 되어 개인정보를 보호하는데 큰 도움을 줄 수 있을 것이라 기대한다.

참고문헌

- [1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster RCNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, Issue 6, pp. 1137-1149, June 2017.
- [2] Jeong Hee Kim, Jung Hun Song, and Dong Hoon Lim, "CT image denoising using inception model," Journal of the Korean Data And Information Science Society, 31(3), pp. 489-500, 2020.