

크로마 키 기법을 적용한 데이터 증강 기법의 효용에 대한 연구

이의재¹, 황큰별¹, 사재학², 박상우²

¹㈜텔로스

²크레텍책임주

euijaebilly@jbnu.ac.kr, telos_inc@naver.com, jhsa@daegu.ac.kr, psw3928@naver.com

Effectiveness of Data Augmentation Using Chroma Key Technique

Eui Jae Lee¹, Keun Byeol Hwang¹, Jae-hak Sa², Sang Woo Park²

¹Telos.Inc

²CRETEC CHEGIM Co., Ltd.

요 약

원본 이미지를 변형하여 학습용 데이터를 확장하는 기법에 대해서는 이전부터 꾸준히 논의된 바가 있다. 턴 테이블과 크로마 키를 이용하여 객체의 영상을 촬영하고 프레임을 추출하여 이미지 분류, 영상 내 객체 탐지 등에 사용이 가능한 데이터 셋의 확장 구축 방안에 대해 다루며, 성능 분석 결과 평균 90% 이상의 객체 검출률을 보였으며 객체 탐지 모델의 경우에서 보다 높은 정확도를 보임을 확인할 수 있었다. 영상내 단일 객체를 인지하기 위한 상황을 위해 본 논문이 제시하는 데이터셋 구축 방안은 충분한 효과를 보일 수 있을 것으로 기대된다.

1. 서 론

컴퓨터 비전 분야의 특수한 결과를 목적으로 하는 모델인 경우 기존의 관련 연구들의 수가 부족하거나 제한된 데이터셋을 이용할 수밖에 없기 때문에 과소적합, 과대적합 등의 문제를 피할 수 없다. 따라서 그에 상응하는 불균형의 이미지 데이터셋의 확보에는 상당한 비용 및 시간 등의 자원이 소모될 필요가 있다.

본 논문에서는 기존 알고리즘 기반 데이터 증강 기법에 크로마 키 기법을 추가 적용하는 데이터 증강 기법으로 이미지 및 영상 모두에서 사용이 가능한 이미지 데이터셋 증강 구축 방안에 대해 연구하고자 한다.

2. 관련 연구

여러 방면으로 다양하게 활용될 수 있는 내용으로 동영상 내 객체 추적을 위한 영상 데이터셋 구축 방법[1]과 본 논문의 유사성을 확인할 수 있었다. 해당 연구에서는 Yolov4 네트워크에 최적화된 실시간 객체 탐지를 위한 데이터셋 구축 방안에 대해 제시하고 있으며 영상 분할에 대한 내용을 주로 참고하였다.

2.1 데이터 증강

2.1.1 데이터 증강 개요

데이터 증강 기법은 기본적으로 데이터의 주요 특징은

유지하고, 이미지를 직간접적으로 변형하여 원본 이미지를 확률적으로 반영한 다른 다수의 이미지를 생성해내는 것을 말한다.

2.1.2 알고리즘 기법

신경망의 네트워크에서 이미지의 픽셀을 벡터의 형태로 학습하는 것으로 원본과 변형된 이미지를 서로 다른 이미지로 인식할 수 있다.

그러나 알고리즘의 변형 기법을 이용한 데이터 증강은 증강기법에 대한 확실한 이해가 필요하다. 즉, 증강된 데이터의 중복, 원본과의 불일치 등으로 오히려 학습효과가 떨어질 수 있다.

2.1.3 딥러닝 기법

GAN은 상당수의 이미지 합성 기술의 기반으로 사용되며 내부적으로 데이터를 합성해내는 생성자(Generator)와 이에 대한 데이터 분포를 판별하는 구분자(Discriminator)가 서로 경쟁하는 구조로 동작한다. 즉, 구분자(Discriminator)는 생성자(Generator)와 기존 확률분포와의 차이를 판단하며, 생성자(Generator)는 원본 데이터의 확률분포에 맞춰 구분자(Discriminator)를 속이기 위해 원본 데이터와의 확률분포 차이를 감소시킬 수 있도록 생성모델을 수정해가며 실제 데이터의 분포에 가장 가까운 데이터를 생성하는 것을 목표로 한다.



(그림 1) cycleGAN 을 이용한 이미지 변환.[2]

2.1.4 전이 학습

전이학습은 학습을 필요로 하는 데이터셋의 크기가 작고 기학습 데이터의 카테고리 이외의 결과를 필요로 하는 경우에 사용한다. 신경망을 초기층과 중간층만으로 특징 맵을 추출하는 추출기로서 사용하여 IMAGENET, COCO 등 범용 데이터 셋을 이용해 대중적인 범위에 대한 기본적인 학습을 수행해 Feature 를 추출한 후 추가로 재학습시켜 사용하는 방식이다. 대부분의 학습 문제를 회피할 수 있으며, 신경망의 성능을 향상시킬 수 있다.

3. 데이터 셋 증강

3.1 영상 촬영

촬영은 크로마 키가 적용된 배경과 턴테이블을 일정한 속도로 회전하여 2가지 이상의 각도로 동영상을 촬영하여 수행한다. 촬영된 영상은 해당 객체의 클래스를 이름으로 하는 디렉토리에 저장하며 파일명에 객체의 이름과 방향, 각도를 명시하였다. 데이터셋으로 구축할 객체를 촬영한 영상으로부터 프레임은 이미지로 추출한다.

3.2 영상 프레임 추출

영상 프레임 추출 기법의 경우 동영상 내 객체 추적을 위한 영상 데이터셋 구축 방법의 영상 분할 내용을 참고하였다. 3.1에서 사용된 턴테이블은 약 30초에 한번 회전하므로 아래 식에 따라 프레임당 약 0.4도로 회전한다는 것을 알 수 있다.

$$R = \frac{360^\circ \times RPM}{30fps \times T}$$

즉, 촬영된 동영상을 10° 단위로 분할하기 위한 분할 단위는 25프레임이 된다.

3.3 이미지 가공 및 라벨링

동영상으로부터 추출된 프레임 이미지에 크로마 키 기법을 적용하여 배경을 제거하였다. 이때 에지(edge)에 그림자 혹은 반사광 등으로 인한 녹색의 잔상이 남게 된다. Green 채널 픽셀의 평균상한을 조정하는 방법으로 에지의 잔상을 보정한다.

배경이 제거되어 컨투어(contour)의 탐지에 용이한 형태를 가지게 되므로 객체의 위치를 특정하는 바운딩 박스(Bounding Box)를 지정하고 해당 객체의 중심을 기준으로 512x512의 영역을 Crop 하여 추출한다. 이미지 분류를 위한 정보의 경우 동영상의 경로에서 이미지 분류에 필요한 클래스 정보를 가져와 저장한다. 생성된 두 종류의 라벨링 정보를 json 과 txt 포맷의 라벨링 데이터로 저장한다.

3.4 이미지 증강 및 정제

데이터 중 경우에 따라 객체가 소실되거나 배경의 제거가 완전하지 않은 경우 등에 해당하는 이미지를 제거하여 데이터를 정제한다. 2.1.2의 알고리즘 기법 중 Rotation, Flip, Scaling의 3가지 기법을 랜덤으로 적용하고 배경이 될 이미지와 합성하여 데이터를 증강한다.

4. 실험

3의 방법대로 데이터셋을 구축하였으나 보다 다양한 각도에서의 이미지를 빠르게 구축하기 위해 촬영장치를 직접 제작하여 영상을 획득하였다.



(그림 2) 제작한 원본 영상 수집 촬영 부스.

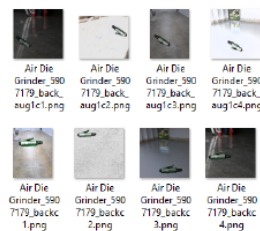
본 실험에서 원본 영상은 25의 분할 단위를 적용하여 10프레임 단위로 영상을 분할하였다. 분할된 데이터셋은 약 4만여 장이 확보되었으며 추가 보정을 거쳐 원본 이미지로 활용하였다.



(그림 3) (좌)크로마 키 적용 이미지 (우)추가 보정 이미지.

원본 이미지는 총 4회의 데이터 증강과 15장의 배경 이미지 중 5장과 랜덤으로 선택 합성 과정을 거쳐 약 80만장의 데이터셋을 확보하였다.

전체 데이터 셋은 train(8):val(1):test(1)의 비율로 분할하고 이미지 분류 성공률을 측정하였다



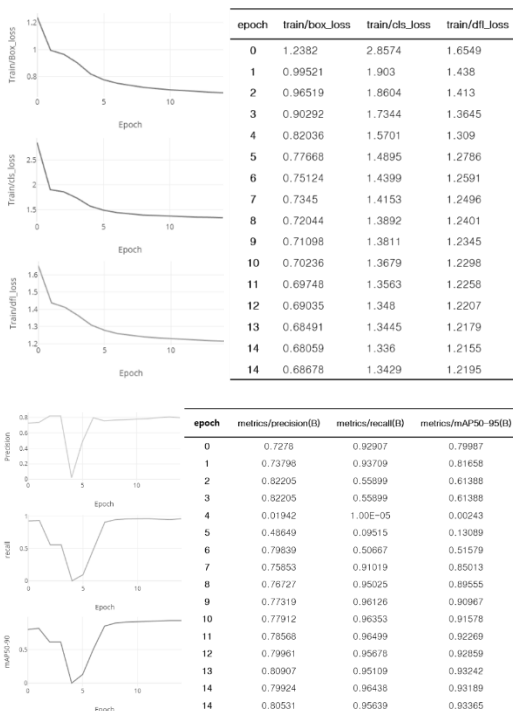
(그림 4) 배경 합성 이미지.

5. 학습 성능 평가

기본적으로 각 모델의 정확도는 학습 진행도에 따라 변동될 수 있어 학습 과정에서 epoch 마다 Validation 및 Accuracy test 를 진행하고 있으며 그래프 또는 표를 이용해 그 결과를 나타내고 있다.



(그림 5) 이미지 분류 모델 학습 그래프



(그림 6) 객체 탐지 모델 학습 그래프

4 epoch 부근에서 높은 오류율을 보였으나 후반 학습이 진행되면서 높은 정확도를 유지하는 모습이 확인되었다.

이와는 별개로 실제 사용환경에서의 정확도를 측정하기 위해 별도의 성능 검증을 진행하였다. 실험 환경의 경우 학습 시 사용한 합성 데이터, 미사용 학습 데이터와 실제 촬영 데이터 3가지 분류로 나눈 약 150여장의 데이터로 진행하였다.

<표 1> 각 모델의 데이터 별 정확도

model	Used	Not used	Others
Resnet10	100.0	88.0	51.334
Yolov8	98.667	86.0	85.334

이미지 분류 모델의 경우 이미지 내에서 객체가 차지하는 범위의 비율이 크지 않아 학습 과정에서 합성된 배경 이미지에 영향을 크게 받는 반면, 객체 탐지 모델의 경우 객체가 차지하는 범위를 특정하여 탐지를 진행하므로 배경 이미지의 영향을 적게 받아 비교적 높은 검출률을 유지하는 것으로 보인다.

5. 결론

다수의 객체가 동시에 존재하거나, 매우 작은 크기의 객체에 대해서는 이미지 데이터 증강을 통한 이미지 분류 모델의 학습량 증가 보다 객체를 명확히 인지하고 탐지하는 객체 탐지 모델이 보다 높은 검출률을 유지할 수 있음을 알 수 있었다.

본 논문에서 진행한 연구를 통해 필요로 하는 사용 환경, 형태의 변화 등의 특징에 대해 크로마 키 기법을 이용하여 한정적인 데이터를 증강함이 부족 데이터의 증강에 있어 비용을 줄이고 정확도를 높일 수 있는 효과적인 대안이 될 수 있을 것으로 보인다. 다만 사용하고자 하는 목적과 특징을 명확히 인지하는 것이 가장 중요할 것으로 생각된다.

6. 참고문헌

[1] J. S. Kim, G. J. Heo and S. W. Jang, " Building Method an Image Dataset for Tracking Objects in a Video "

[2] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2223-2232.