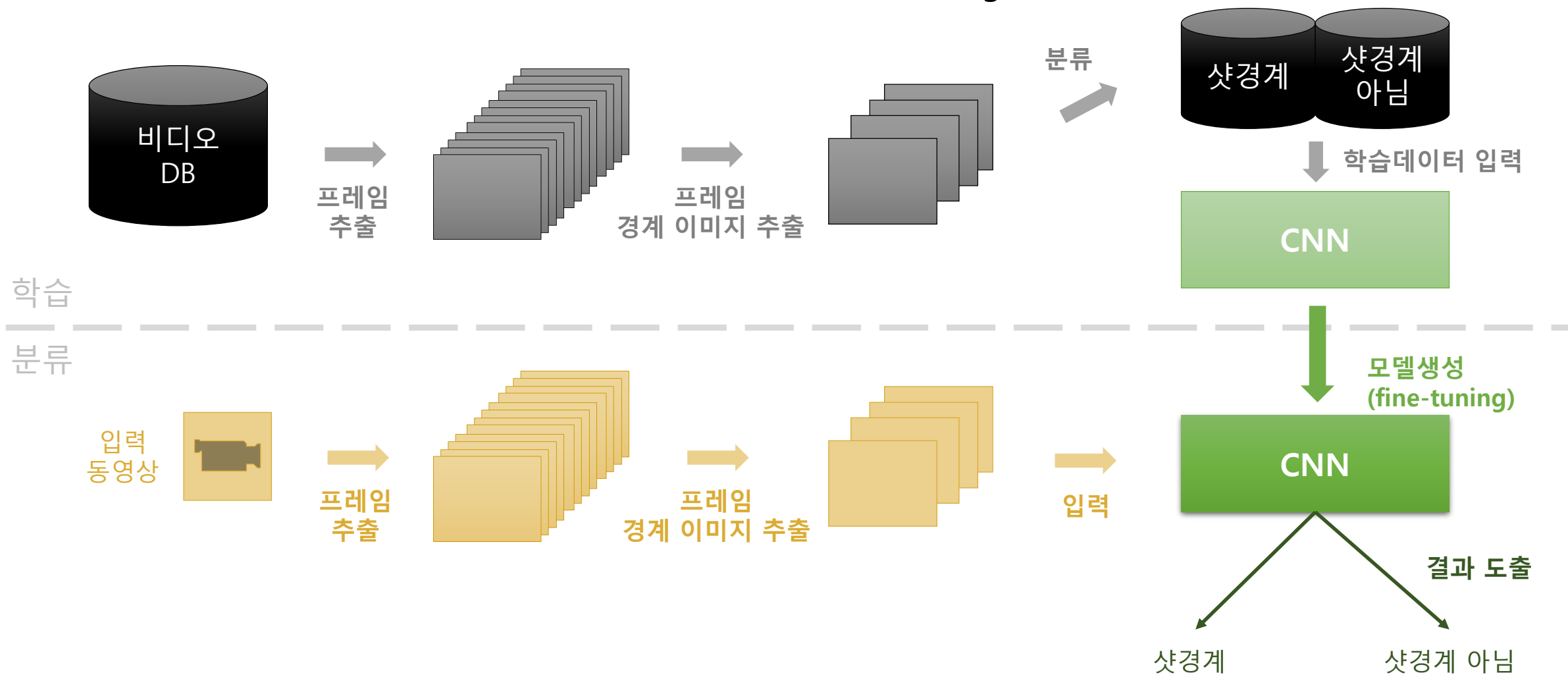


MobileNetV2 Fine-tuning을 통한 Shot boundary classification 구현

신경회로망특론 Term Project
120190198 석혜경

CNN을 이용한 샷 경계 검출

- 본 프로젝트는 CNN을 이용한 샷 경계 검출 방법을 제안한다.
- CNN을 이용한 샷 경계 검출 모델은 자체 데이터셋으로 fine-tuning 된 2-label classification 모델이다.

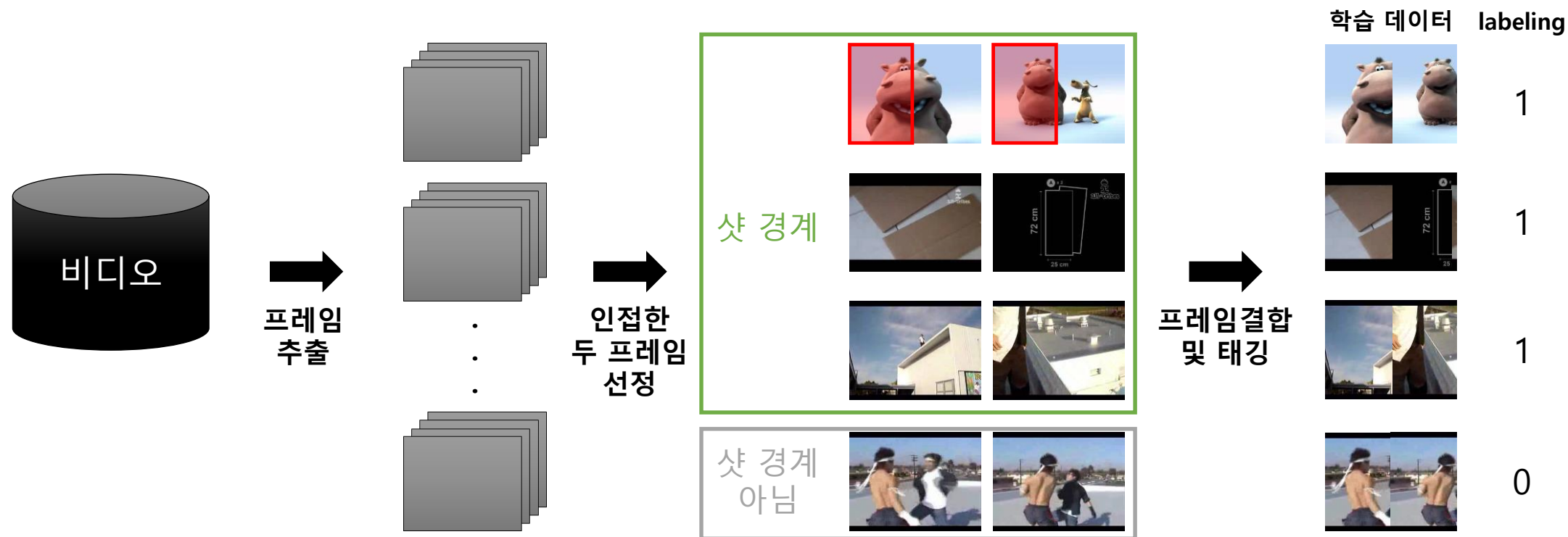


MobileNetV2 (CNN model)

- 2017년에 발표된 MobileNet에 이어 2019년에 발표된 경량화 네트워크이다.
- 컴퓨터 성능이 제한된 모바일, 디바이스 등에서 사용될 목적으로 설계되었다.
- Channel Reduction, Depthwise Separable Convolutions, Distillation & Compression 기법을 통해 small deep neural network를 구현하였다.
- 논문에서 제안된 Inverted Residuals 구조와 Linear Bottleneck 도입으로 기존 MobileNet에서 발전시킨 MobileNetV2가 발표되었다.
- PyTorch를 사용하여 pre-trained MobileNetV2 모델을 fine-tuning 하였다.

경계 이미지 구성

- shot boundary detection을 위한 데이터셋 재구성을 위해 샷 경계에 해당하는 두 프레임을 뽑고, 두 프레임을 결합하여 한 프레임으로 나타낸다.
- 비디오는 CC_WEB_VIDEO, TRECVID2018 일부를 포함한다.



경계 이미지 예시



샷 경계인 경계이미지



샷 경계가 아닌 경계이미지

실험

- train set, validation set은 CC_WEB_VIDEO 데이터셋과 TRECVID2018 데이터셋의 일부인 35개의 비디오로부터 추출한 경계 프레임 이미지로 구성되었다.
- test set은 VCDB core 데이터셋 28개의 비디오로부터 추출한 경계 프레임 이미지로 구성되었다.
- 이미지의 중심을 기준으로 256 x 256으로 이미지를 잘라서 사용한다.

	train	validation
샷 경계	847	193
샷 경계 아님	868	194
합계	1715	387

학습 데이터 개수
(CC_WEB + TRECVID2018)

	test
샷 경계	749
샷 경계 아님	798
합계	1547

테스트 데이터 개수
(VCDB_Core)

결과

		실제 정답	
		True	False
분류결과	True	628	17
	False	121	781

confusion matrix

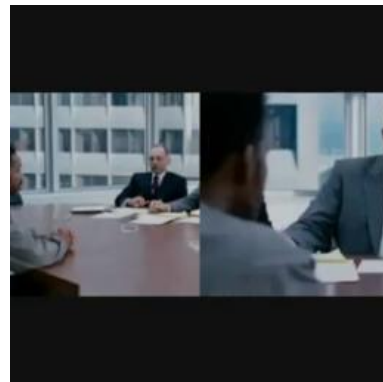
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 0.91$$

$$Precision = \frac{TP}{TP + FP} = 0.97$$

$$Recall = \frac{TP}{TP + FN} = 0.84$$

결과 분석

- 잘못 분류한 이미지의 경우,
black border가 다른 이미지에 비해 크게 들어가거나 같은 로고가 삽입된 경우가 많다.



향후 연구

- 학습 데이터셋을 늘려 재학습을 시도한다면 성능 향상을 기대할 수 있다.
- 성능을 낮추는 이미지 내 black border를 제거 및 처리하여 테스트하거나 black border를 포함하는 경계 이미지 데이터를 학습 데이터에 추가하여 학습을 시도할 수 있다.
- fade나 dissolve 등 두 인접 프레임만으로는 구분이 잘 가지 않는 샷 경계의 경우를 고려한다.