

# 10명의 손글씨 필적감정 CNN 모델



한국산업기술대학교 IT경영학과  
최지아, 하성민  
2019. 12. 10



## 프로젝트 발표 목차

---

1. 주제 변경
2. 데이터 수집
3. 데이터 전처리
4. 모델 구성
5. 성능 향상 시도



## 1. 주제 변경

### I. 기존 주제 : 문서 내 수기 서명 위/변조 판별 모델

- 데이터의 수 부족( $14\text{명} \times 40\text{개} = 560\text{개}$ ), 학습이 어려울 것 같다는 판단
- 가짜 데이터 수집의 어려움



## 1. 주제 변경

### II. 변경 주제 : 10명의 손글씨 필적감정 CNN 모델

- 선정이유 : 서명위조와 유사하게 사람의 특성을 파악 & 비교적 많은 데이터 확보 가능
- 필적 감정의 문제점 존재



① 과도한 시간 소요  
(공책 하나 당 5일 소요)



② 과도한 비용 발생  
(공책 하나 당 50만원)



③ 객관성 결여  
(감정원 마다 다른 감정)

출처:한국문서감정사협회



## 2. 데이터 수집

가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가
가	나	나	나	나	나	나	나	나
가	나	나	나	나	나	나	나	나
가	나	나	나	나	나	나	나	나
가	나	나	나	나	나	나	나	나
가	나	나	나	나	나	나	나	나
가	나	나	나	나	나	나	나	나
가	나	나	나	나	나	나	나	나

- 10명 필체 수집
- 6가지의 서로 다른 글자  
(가, 나, 다, 라, 마, 바)
- 한 사람당 한 글자에 대해 63회 기입
- 총 3780개의 데이터 수집



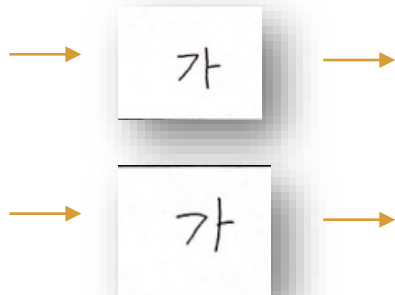
### 3. 데이터 전처리(OpenCV)

#### 1. 기존 데이터

가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가

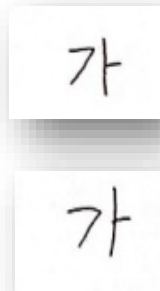
- Image resize  
(width : 1000pixel 고정)

#### 2. 이미지 분할



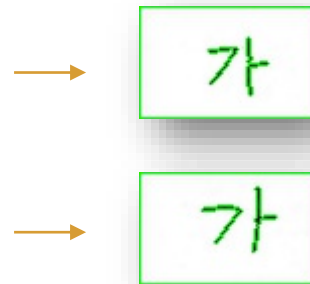
- OpenCV Image crop  
(비율에 맞춰 cropping)

#### 3. 두번째 이미지 분할



- 첫번째 이미지 분할 시  
잘못하지 못했던 부분 처리

#### 4. 글자 데이터 윤곽 찾기



- OpenCV contour  
(이미지 데이터를  
더 명확하게 표현)



### 3. 데이터 전처리

#### 5. 데이터 분할

test	2019-11-23 오후 6:53
train	2019-11-23 오후 6:53
valid	2019-11-23 오후 6:53

이름	수정된 날짜	유형
1	2019-11-23 오후 6:54	파일 폴더
2	2019-11-23 오후 6:54	파일 폴더
3	2019-11-23 오후 6:54	파일 폴더
4	2019-11-23 오후 6:54	파일 폴더
5	2019-11-23 오후 6:54	파일 폴더
6	2019-11-23 오후 6:54	파일 폴더
7	2019-11-23 오후 6:54	파일 폴더
8	2019-11-23 오후 6:55	파일 폴더
9	2019-11-23 오후 6:55	파일 폴더
10	2019-11-23 오후 6:55	파일 폴더

- (가)~(바)까지의 글자들이 동일한 비율로  
훈련셋/검증셋/시험셋을 구성할 수 있도록 데이터 분할





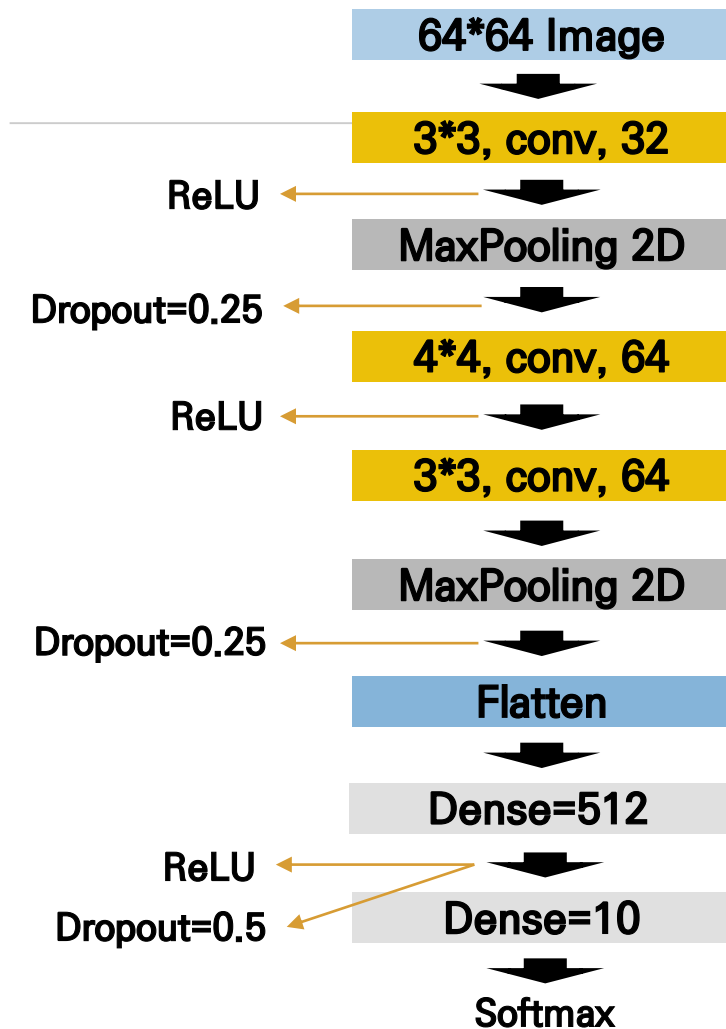




## 4. 모델 구성

### 1. 모델 구조

- Convolution Layer 3개
- MaxPooling Layer 2개
- Flatten Layer 1개
- Dense Layer 2개
- Dropout 사용
- 활성화함수 : ReLU
- 마지막 출력층 활성화함수 : Softmax





## 4. 모델 구성

### 2. 추가 구성

- $X_{\text{값}}(X_{\text{train}}/X_{\text{valid}}/X_{\text{test}})$   
→ Conv2D 적용 위해 reshape(64\*64)
- $Y_{\text{값}}(Y_{\text{train}}/Y_{\text{valid}}/Y_{\text{test}})$   
→ One-hot encoding
- Early Stopping 적용

### 3. 모델 학습 파라미터 구성

- Optimizer : adam → 최적의 Optimizer 찾기 위한 시도 →
- Epochs : 200
- Batch\_size : 20

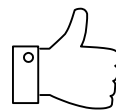
rmsprop : 85.28%

nadam : 10.00%

adagrad : 10.00%

adam : 85.83%

adadelta : 82.64%



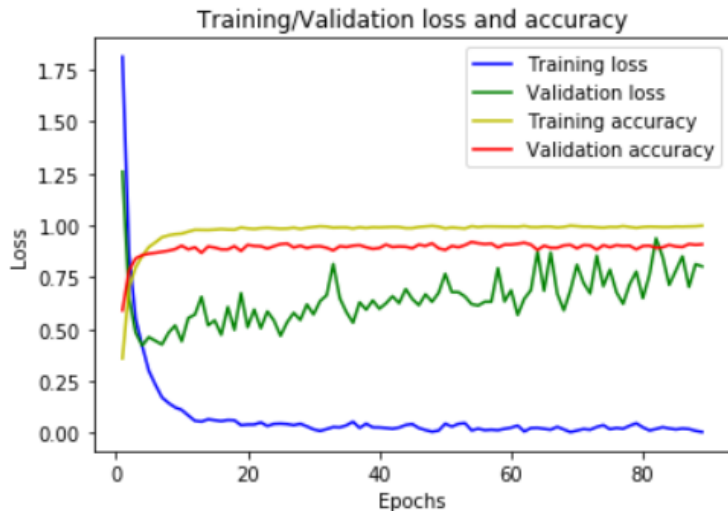


## 4. 모델 구성

### 4. 모델 학습 결과

- Epochs : 200  
→ (Early Stopping) Epochs : 77
- Batch Size : 20
- Accuracy : 85.83%

acc : 85.83%





## 5. 성능 향상 시도

### 1. Data Augmentation

- Rotation / Zoom / 가로 이동 / 세로 이동 → 무작위 적용
- 1개의 이미지 데이터를 4개로 변환 (기존 데이터 수 \* 4)
- 훈련셋 : 9120개, 검증셋 : 3120개, 시험셋 : 2880개  
= 총 15,120개

#### ① Contour를 하지 않은 이미지



#### ② Contour를 한 이미지



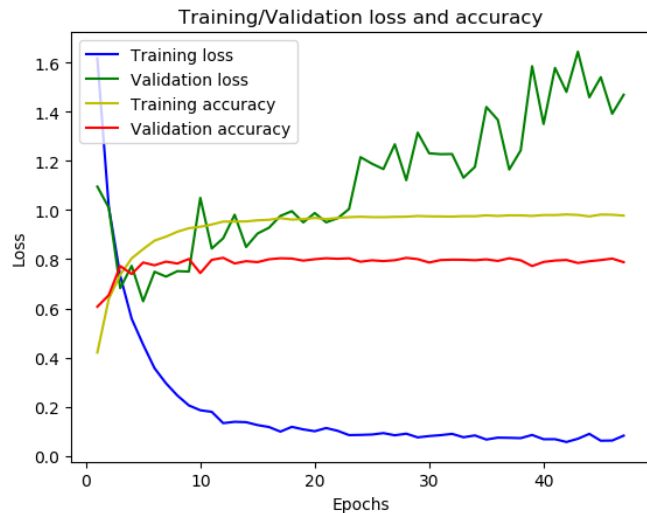


## 5. 성능 향상 시도

### 2. 모델 학습 결과

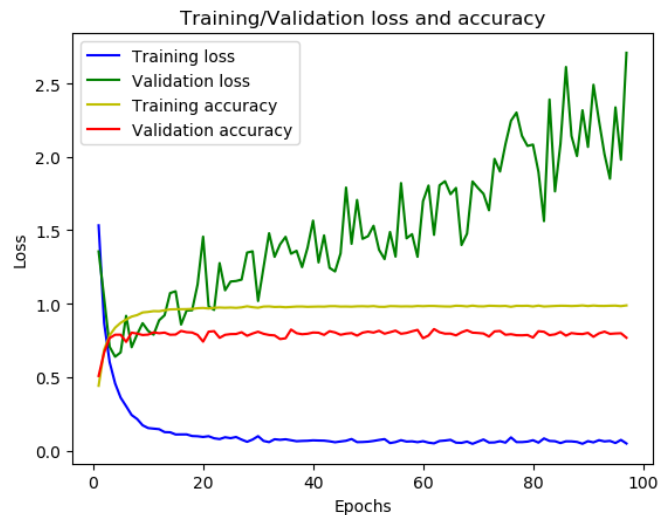
#### ① Contour를 하지 않은 이미지

- Accuracy : 76.22% (성능 하락..)



#### ② Contour를 한 이미지

- Accuracy : 71.56% (성능 하락..)





# Q&A

---