

필적감정을 위한 기계학습 모델 구 축 및 학습데이터 생성 방법

제40회 한국법과학회 2021 춘계학술대회

김재우, 한국전자통신연구원

개요(Introduction)

필적 감정

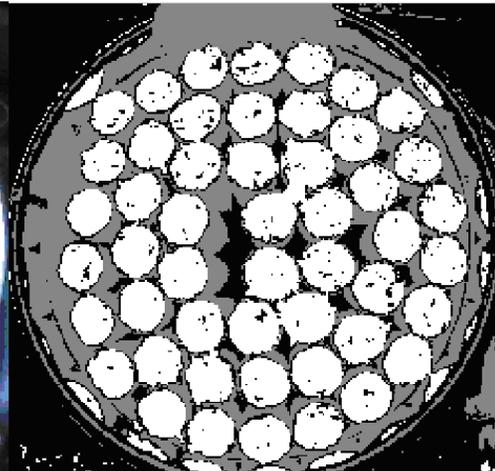
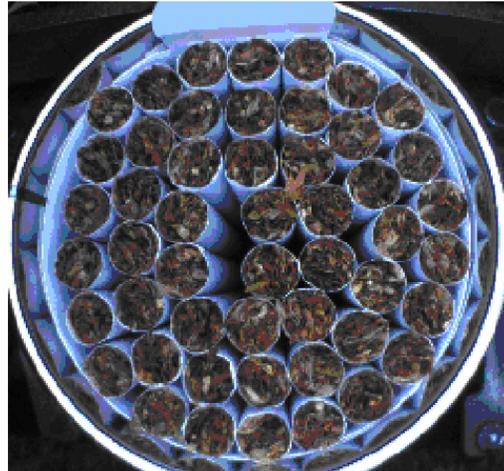
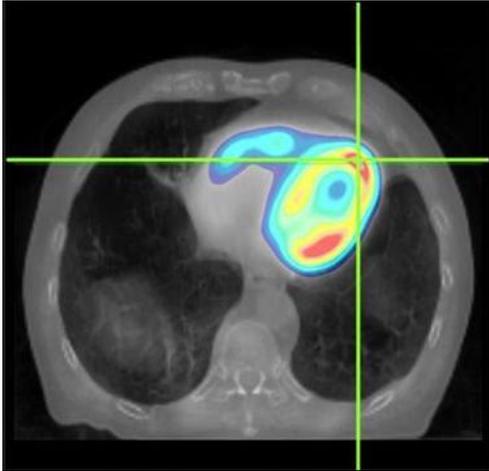
- 두 필기 문장 또는 텍스트가 동일인에 의해 작성되었는지 판별
- 대필/위조 여부 검증, 서명 검증, 범죄 수사 등에 활용

필적 감정 작업의 특징

- 필적은 큰 변화성을 지님
 - 특정 필자의 필적에서 나타나는 공통된 특성(항상성)
 - 특정 필자의 필적에서만 반복적으로 나타나는 특성(희귀성)
 - 숙련된 필적 감정관들의 축적된 지식, 기술, 노하우에 의해 수행됨
 - 오랜 기간의 훈련이 필요
 - **감정 오류의 가능성 존재** →
 - 필적 감정의 정확도 향상을 위해:
 - 객관적인 분석 방법이 필요함
 - 필적감정 지원 시스템이 필요함
- Computer Vision
 - Pattern Recognition
 - Deep Learning

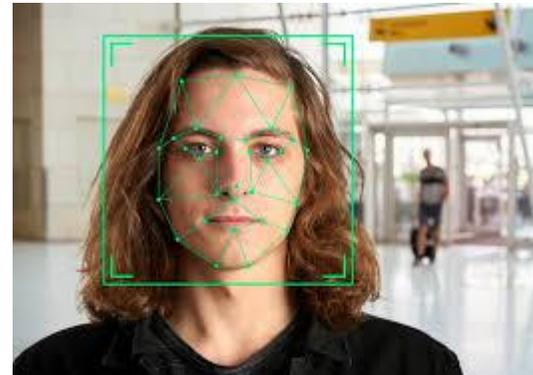
컴퓨터 비전 기술 개요

영상처리, 컴퓨터 비전, 패턴 인식, 기계학습, 딥러닝, ...

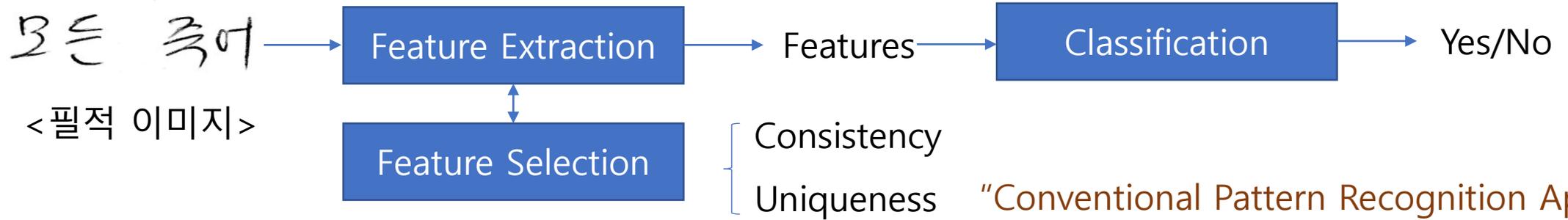


```
101010111001  
000101001000  
100100111000  
101010101010  
100100110010  
010010101010  
101001101101
```

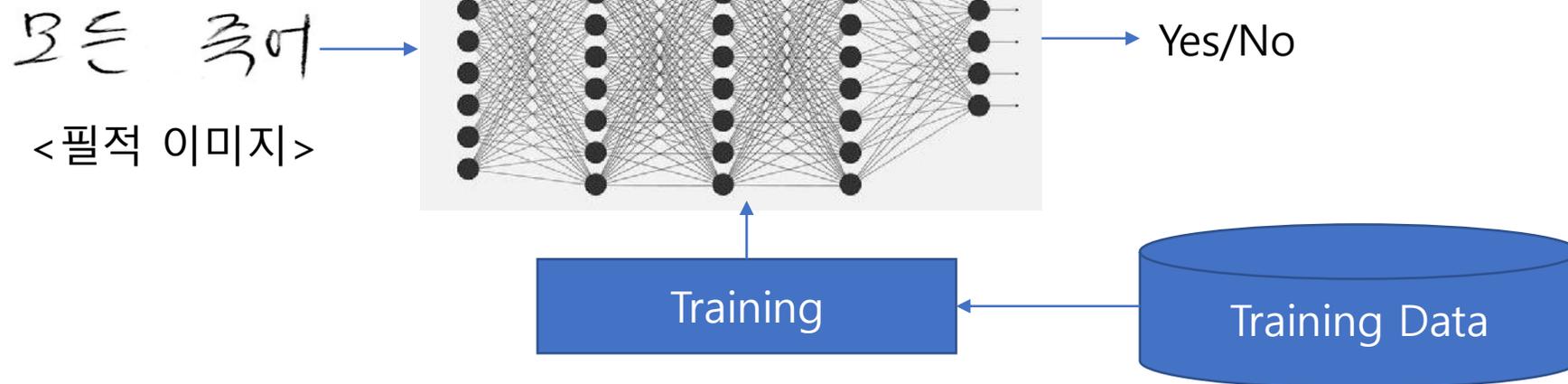
Digital template of the fingerprint



필적 감정을 위한 컴퓨터 비전 기술



“Deep Learning Approach”



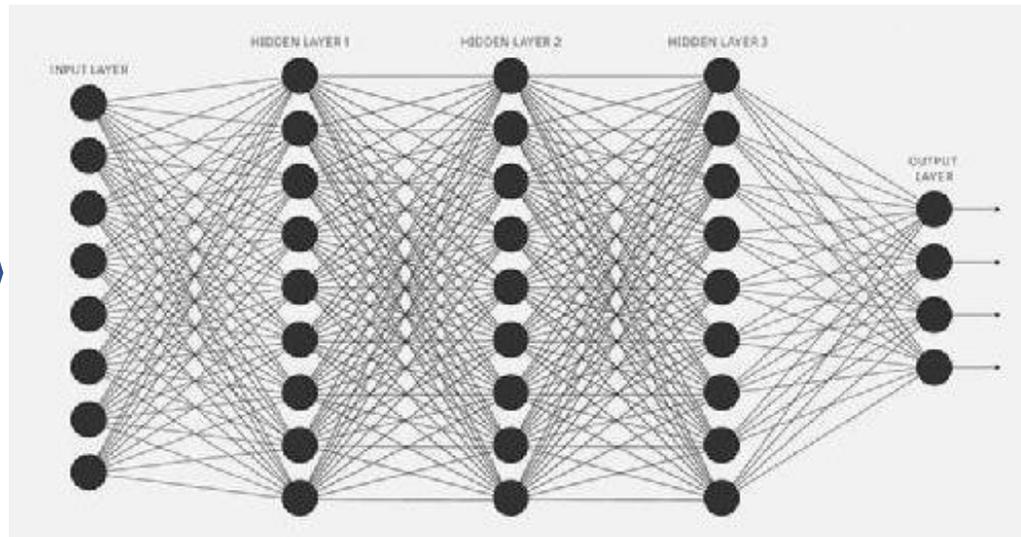
컴퓨터 비전 + 기계학습



<Cats>



<Dogs>

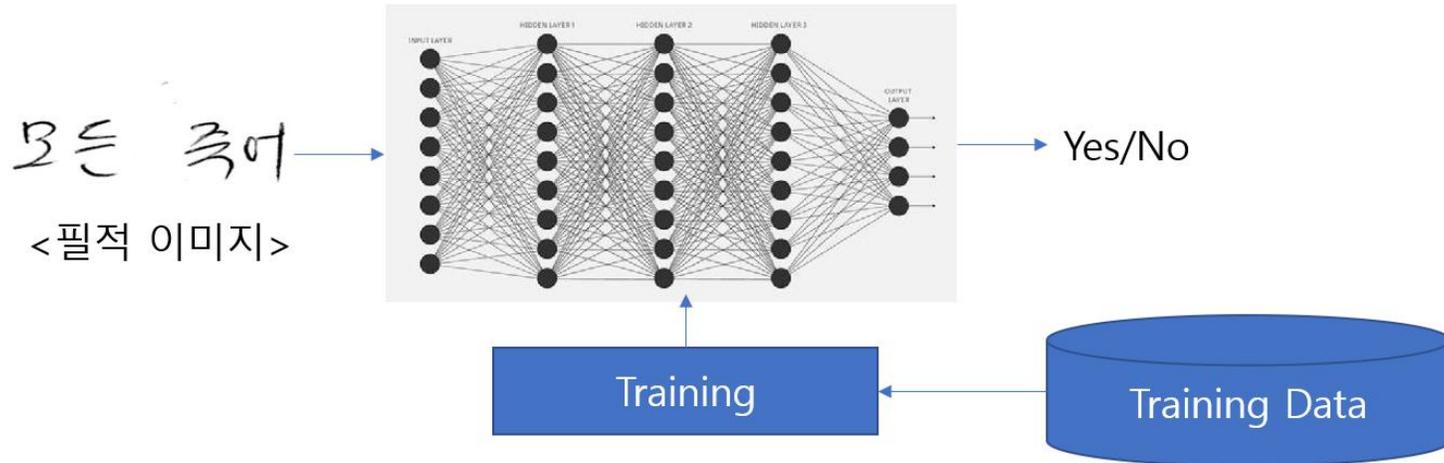


Cat/Dog

딥러닝 학습 모델의 정확도는
학습데이터의 양과 질이 좌우함.

딥러닝을 위한 학습 데이터 생성

“Deep Learning Approach”



죽는 날까지 하늘을 우러러 한 점 부끄럼이 없기를,
 잎새에 이는 바람에도 나는 괴로워 했다.
 별을 노래하는 마음으로 모든 죽어 가는 것을 사랑해야지
 그리고 나한테 주어진 길을 걸어가야겠다.
 오늘 밤에도 별이 바람에 스치운다.

학습 데이터 자동생성이 필요



220명의 필적 샘플, 총 5,504 페이지 분량

연구 목표

- 필적 감정용 딥러닝 학습 모델을 위한 학습 데이터 자동 생성
- 오류 발생 시 사용자 개입에 의한 보정 작업의 최소화
- 사전에 알고 있는 파라미터들 이용 자동 생성 정확도 향상

감정용 필기 데이터 생성

죽는 날까지 하늘을 우러러 한 점 부끄럼이 없기를,
앞새에 이는 바람에도 나는 괴로워했다.
별을 노래하는 마음으로 모든 죽어 가는 것을 사랑해야지
그리고 나한테 주어진 길을 걸어가야겠다.
오늘 밤에도 별이 바람에 스치운다.

시필 내용

- 사전 정의된 구조
- → 5줄로 구성
- → 각 줄당 8, 5, 8, 5, 5 단어가 존재
- 단어 분리/레이블링시 자동화에 도움

죽는 날까지 하늘을 우러러 한 점 부끄럼이 없기를,
앞새에 이는 바람에도 나는 괴로워했다.
별을 노래하는 마음으로 모든 죽어 가는 것을 사랑해야지
그리고 나한테 주어진 길을 걸어가야겠다.
오늘 밤에도 별이 바람에 스치운다.

죽는 날까지 하늘을 우러러 한 점 부끄럼이 없기를,
앞새에 이는 바람에도 나는 괴로워했다.
별을 노래하는 마음으로 모든 죽어 가는 것을 사랑해야지
그리고 나한테 주어진 길을 걸어가야겠다.
오늘 밤에도 별이 바람에 스치운다.

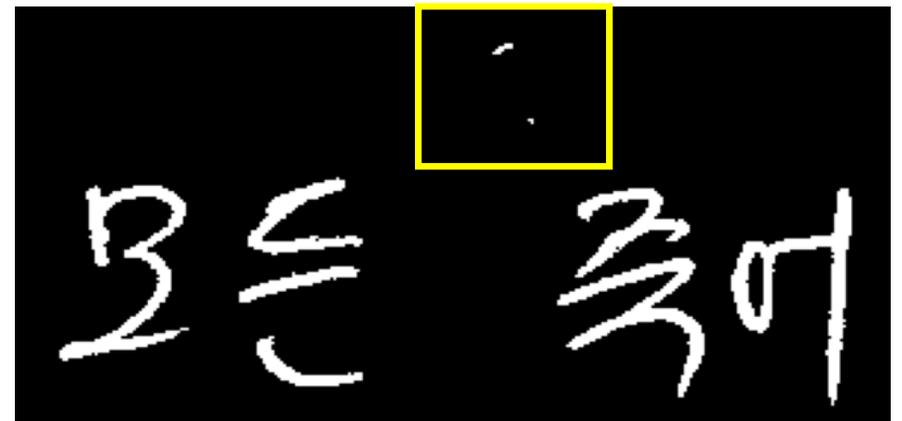
필기 스캔 이미지 데이터

죽는 날까지 하늘을 우러러 한 점 부끄럼이 없기를,
앞새에 이는 바람에도 나는 괴로워 했다.
별을 노래하는 마음으로 모든 죽어 가는 것을 사랑해야지
그리고 나한테 주어진 길을 걸어가야겠다.
오늘 밤에도 별이 바람에 스치운다.

모든 죽어

잡음 제거(noise reduction)

모든 죽어 $\xrightarrow[\text{Threshoding}]{\text{잡음제거}}$ 모든 죽어

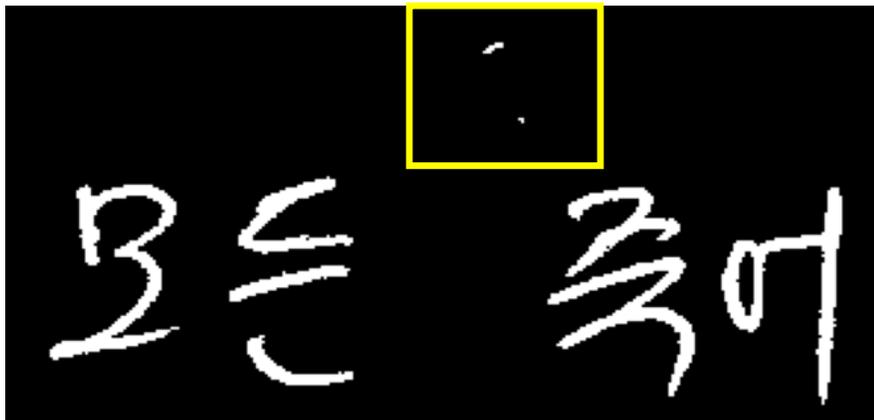


잡음 제거(noise reduction)

연결구성요소의 크기(픽셀 개수)가 일정크기 이하이면 잡음으로 판단함.

모든 죽어

연결구성요소 검출
→
(Connected components)



모든 죽어



라인 검출-모폴로지 연산

모폴로지 연산, 팽창(dilatation) 연산

- 자모, 글자, 단어 검출 시 많이 사용하는 방법
- 전경을 주변으로 확장하는 연산
- 사전 정의된 윈도우
- 중심 픽셀이 전경, 윈도우 내 모든 픽셀을 전경으로



수평방향, 20픽셀 확장

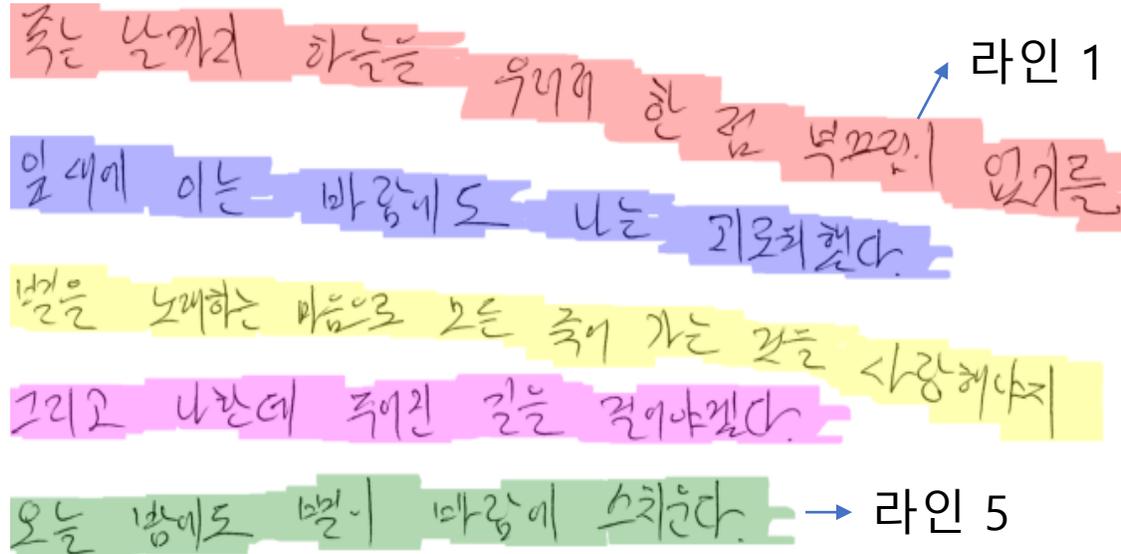


수평방향, 40픽셀 확장



수평/수직 방향, 30픽셀 확장

라인 검출



- 확장된 전경 이미지로 부터 연결구성요소를 라인으로 판단
- 사전 정의된 라인 수, 라인당 단어 수 고려
윈도우 크기 설정
- ➔ 라인 검출 정확도 향상

Algorithm line_detection:

Input: Foreground image

Output: Line labels for each pixel

begin

Initialize window size;

number_of_lines = -1;

Set **predefined_number_of_lines**;

while(**number_of_lines** != **predefined_number_of_lines**)

begin

Adjust window size;

Apply dilatation on foreground image;

Obtain connected components and label the lines;

Get **the number_of_lines**;

end

end

단어 검출

죽은 날까지 하늘을 우러러 한 걸음 부끄럼 없이
앞에 이는 바람에도 나는 괴로워했다. 잊기
별을 노래하는 마음으로 모든 죽어 가는 것을
그리고 나란히 죽어간 길을 걸어야 했다. <사랑해나>
오늘 밤에도 별이 바람에 스치운다.

Algorithm word_detection:

Input: Foreground image segment for a line

Output: Word labels for each pixel

begin

Initialize window size;

number_of_words = -1;

Set **predefined_number_of_words**;

while(**number_of_words** != **predefined_number_of_words**)

begin

Adjust window size;

Apply dilatation on foreground image;

Obtain connected components and label the words;

Get **the number_of_words**;

end

end

단어 추출

← 109_2

필적 감정 대상:	0
비정상/비서명 단어:	0
정상/서명 단어:	191
아이템 소유 페이지:	-

아이템 레이아웃
 정렬 없음
 크기 우선
 텍스트 우선

→

이상의

<내용에는>
내용에는 내용에는 내용에는 내용에는 내용에는 내용에는 내용에는 내용에는
내용에는 내용에는 내용에는 내용에는

<조금도>
조금도 조금도

<거짓이>
거짓이
거짓이 거짓이

<없음을>
없음을
없음을

각 단어의 레이블로 부터 단어 영역을 설정한 후 단어 이미지를 추출함.

결과 및 토의

실험 데이터

- 220명의 필적 샘플
- 총 5,504 페이지 분량
- 9,923 x 14,040 픽셀 (1,200dpi, 21 x 29.7cm)

학 인 지

본인은 2014년 8월 하순경 중앙선법에 근무하는 김민정의 요청으로

일금 사십구억 오천팔백칠십이천오백삼십육원 (₩ 4,958,725,613)에 대한

물류 거래 내역서와 대금 영수증을 무연물산 명의로 작성하여 발송하였나.

그러나 위의 물세가 위법한 관행으로 개인이 물임을 알지 못하여

물류 거래 내역서와 대금 영수증은 1회씩 작성하여 제공하였음 4만

대만달러 동량으로 8월 21일 (₩ 80,000,000)을 입금 받은 것임

위의 내역과 무연물산 명의로 작성하여 발송한 것임

본인이나 제삼자가 작성한 것이 아니라 위법한 관행으로 제공된 것임

서경림입니다.

실험결과

019 - 1

본인필적 1

(관문)

이상의 이상위
내용에는 내용이 거짓인
조금도 거짓이
없음을
확인합니다.
위 내용이 거짓인 경우에 어떤 처벌도 감수하겠습니다.
감수하겠습니다.

- 필적 감정 대상 여부
- 단어 상태: 비정상 또는 비서명
- 단어 상태: 정상 또는 서명
- 단어 상태: 판단 불가
- 앞으로 가져오기
- 뒤로 보내기
- 단어 쪼개기

성명: 권형진

주소: 충청북도 청주시 흥덕구 월명로 848-59 (봉명동)

성명: 권형진

주소: 충청북도 청주시 흥덕구 월명로 848-59 (봉명동)

연락처: 010-9486-5893

연락처: 010-9486-5893

실험결과

← 109_2 →

필적 감정 대상:	0	아이템 레이아웃
비정상/비서명 단어:	0	<input type="radio"/> 정렬 없음
정상/서명 단어:	191	<input checked="" type="radio"/> 크기 우선
아이템 소유 페이지:	-	<input type="radio"/> 텍스트 우선

<본인 필적>

어떤 거것이 위 거것인 조금도 경우에 위 처벌도 거것인 조금도 거것인
 조금도 위 거것이 위 어떤 거것이 내용에는 내용이 이상의 이상의 내용에는
 처벌도 이상의 처벌도 어떤 경우에 경우에 어떤 감소하겠습니다. 내용이
 처벌도 경우에 거것인 어떤 처벌도 내용에는 위 어떤 내용이 없음을 거것이
 없음을 경우에 거것이 연락처 : 010 - 9486 - 5893 없음을 조금도 거것인
 처벌도 위 조금도 경우에 조금도 거것인 설명 : 권형건 위 거것이 어떤
 거것이 경우에 감소하겠습니다. 처벌도 이상의 위 이상의 내용이 거것이 없음을
 내용이 내용에는 연락처 : 010 - 9486 - 5893 거것인 어떤 감소하겠습니다.
 조금도 내용이 감소하겠습니다. 거것인 처벌도 조금도 이상의 감소하겠습니다.
 연락처 : 010 - 9486 - 5893 경우에 내용에는 없음을 내용이 내용에는
 확인합니다. 확인합니다. 이상의 이상의 내용이 확인합니다. 내용에는

← 109_2 →

필적 감정 대상:	0	아이템 레이아웃
비정상/비서명 단어:	0	<input type="radio"/> 정렬 없음
정상/서명 단어:	191	<input checked="" type="radio"/> 크기 우선
아이템 소유 페이지:	-	<input type="radio"/> 텍스트 우선

이상의

<내용에는>

내용에는 내용에는 내용에는 내용에는 내용에는 내용에는 내용에는 내용에는
 내용에는 내용에는 내용에는 내용에는

<조금도>

조금도 조금도

<거것이>

거것이
 거것이 거것이

<없음을>

없음을 없음을 없음을 없음을 없음을 없음을 없음을 없음을 없음을
 없음을

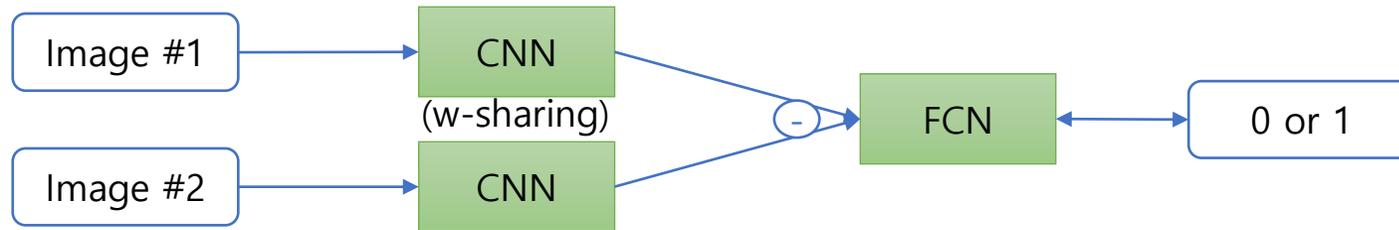
결론

- 필적 감정용 딥러닝 학습 모델을 위한 학습 데이터 자동 생성
- 학습 데이터 생성용 소프트웨어 도구 개발
- 많은 양의 데이터를 빠르고 정확하게 학습 데이터로 변환
- 딥러닝 모델 구축 및 학습 진행중

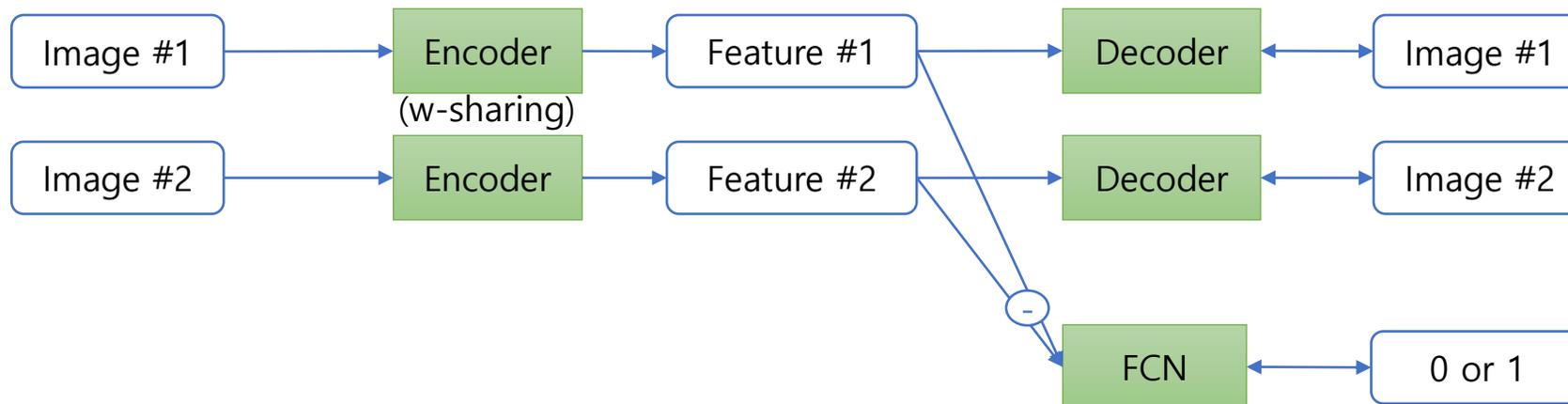
Solution model

보충설명

단순 CNN feature learning (baseline)



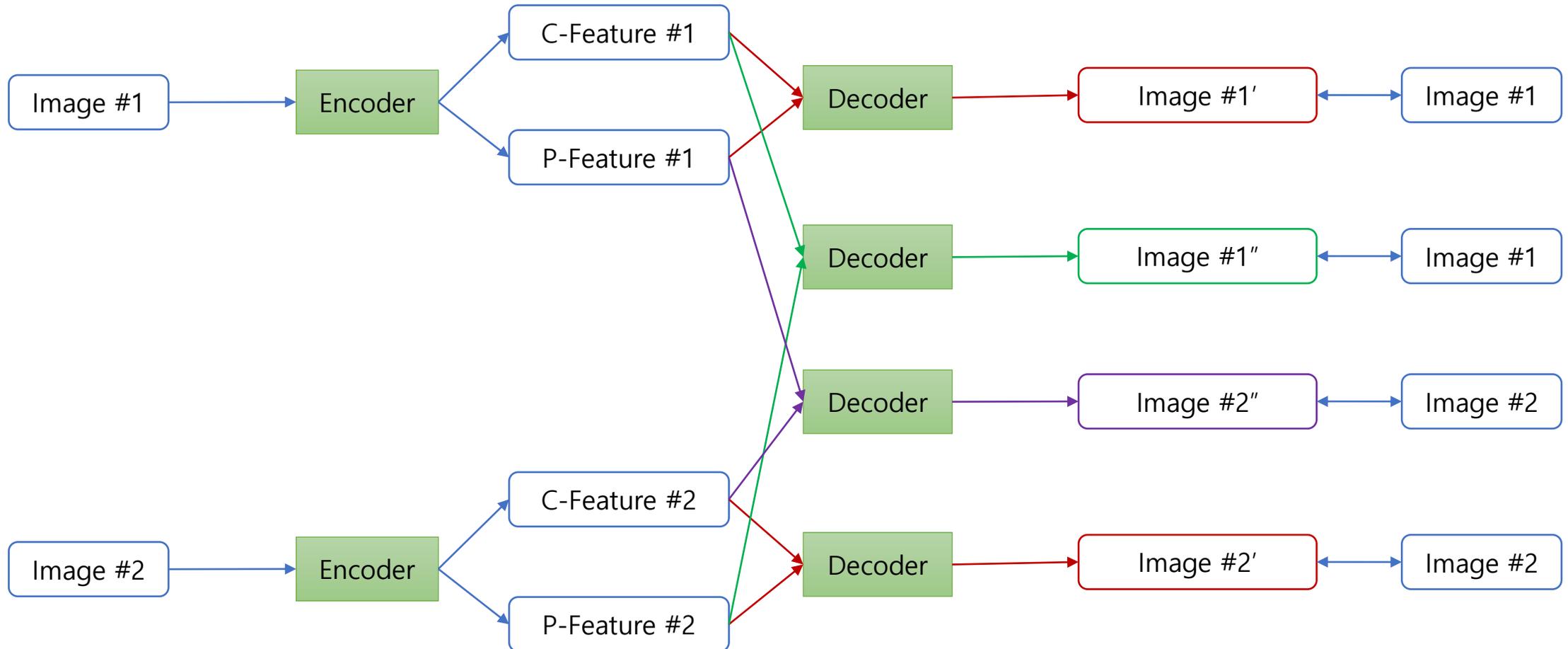
단순 VAE feature learning (baseline)



Solution model

보충설명

- Content/Personality 구분 VAE Feature Learning 아이디어



* Image #1과 #2는 동일 필자의 단어 이미지