

SAMSUNG 삼성증권 DN3A 데모데이 발표

인공지능 OCR 기반 문서 자동 분류

2021-11-22

COMTRUE[®]
Technologies

(주)컴트루테크놀로지

홈페이지 : <http://www.comtrue.com> | 인공지능 서비스 브랜드 : <https://cometrue.ai> | 인공지능 API 서비스 : <https://cometrue.ai/index.php/api-center/>



PoC 개요

목표

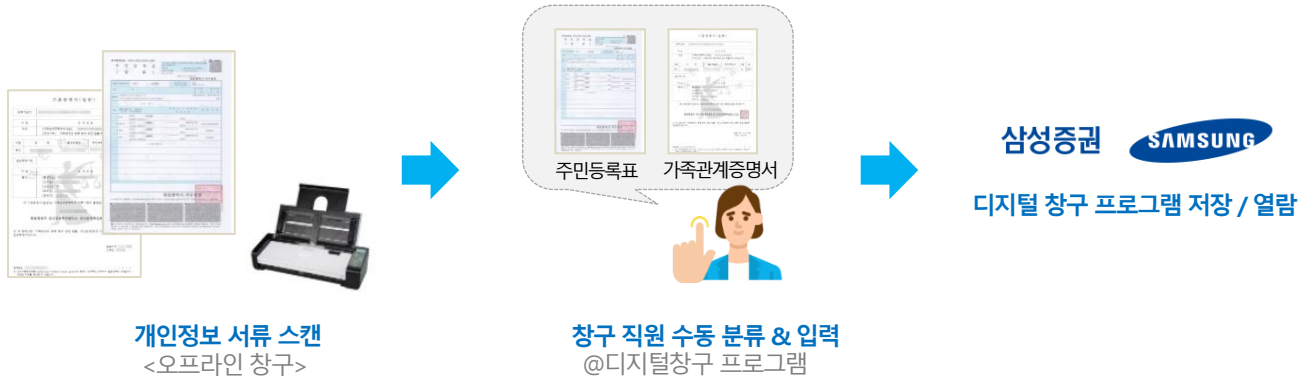
→ 프로젝트 목표

- 디지털 창구 프로젝트 내 **개인정보 문서 자동 분류** 수행

→ 현업 Needs

- 현 디지털 창구 프로젝트 구현에 있어서 창구 스캔 문서에 대한 문서 자동 분류 적용
- CSPB 업무 간소화 및 문서분류 착오방지

As Is



PoC 결과

추진 내용 - 적용 기술 - 인공지능 OCR + 사물인식 응용

신분증 분류



주민등록증

운전면허증

여권

외국인등록증

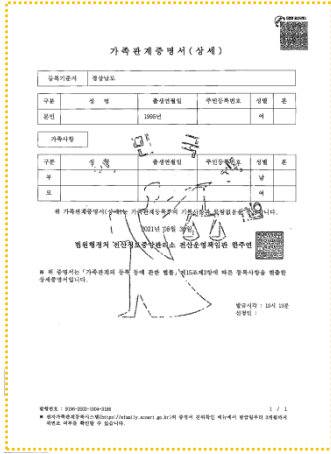
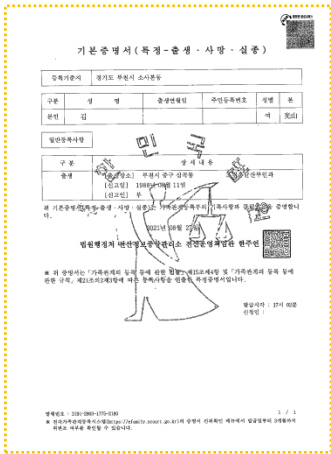


- 신분증 OCR 신경망 기술 노하우 적용
- 신분증 4종에 대한 모바일 및 스캔(흑백, 200dpi, tiff) 문서 인식 및 분류 기술 개발

PoC 결과

추진 내용 - 적용 기술 - 인공지능 OCR + 사물인식 응용

문서분류 OCR 고도화

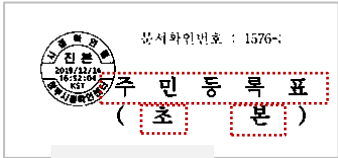


기본증명서 (특정-출생·사망·실종)

기본증명서 특정 출생 사망 실종

가족관계증명서 (상세)

가족관계증명서 상세



주민등록표 초본



기타

- 문서 데이터 수집 및 Oversampling
- 문서 내 이미지 Feature 기반 딥러닝 수행
- 해당 환경에서의 문자인식 커스터마이징(글영역, OCR) 및 고도화

PoC 결과

Windows Library 구현 - 스캔 스테이션, 디지털 창구 연동

To Be



개인정보 서류 스캔
<오프라인 창구>

The software interface displays a document titled '기본증명서' and a classification table. The table lists document numbers and their classification times. A blue bar at the bottom highlights '7 기본증명서' with a classification time of 401ms. The ai see CLASSIFY logo is visible at the bottom right.

No.	문서명	분류시간
1	가족관계증명서	239ms
2	가족관계증명서	307ms
3	가족관계증명서	203ms
4	가족관계증명서	312ms
5	기본증명서	302ms
6	기본증명서	298ms
7	기본증명서	401ms
8	기본증명서	351ms
9	입양관계증명서	331ms
10	입양관계증명서	302ms
11	주민등록표	136ms
12	주민등록표	166ms
13	출입국에 관한 사실증명서	189ms
14	친양자입양관계증명서	377ms
15	친양자입양관계증명서	237ms
16	혼인관계증명서	302ms
17	혼인관계증명서	310ms
18	입양관계증명서	195ms
19	계정입양증명서	178ms

문서분류 신경망 연동 API에서 분류 수행
- 디지털 창구 프로그램에 자동 입력 -



삼성증권 SAMSUNG
디지털 창구 프로그램 저장 / 열람

시연

PoC 결과

결과물 구현 정도

➔ 문서 분류 라이브러리

- 25종 개인정보 문서분류
- 최종 정확도 98% 이상
- 업데이트 DLL 사이즈 ~40MB 이내
- 인식속도 : 0.1~0.5초 이내

총 25개 CLASS (기타 포함) 분류 수행 성공

1차 PoC (13)

신분증	주민등록증 (스캔) 운전면허증 (스캔)
공문서 (개인)	가족관계증명서 기본증명서 혼인관계증명서 입양관계증명서 친양자입양관계증명서 주민등록표 초본 개인인감증명서 출입국에관한 사실증명서
공문서 (기업)	사업자등록증 법인인감증명서 법인등기부등본

2차 (11)

신분증	외국인등록증 (스캔) 여권 (스캔)
공문서 (개인)	본인서명사실확인서 건강보험증 근로소득원천징수영수증 후견등기사항증명서 소득확인증명서(ISA가입용) 소득금액증명원 거주자증명서 은행통장사본
공문서 (기업)	사업자등록증명원
기타	기타분류

기대효과

삼성증권



➔ 업무 생산성 향상

- 디지털 창구 문서 분류 자동화
- 수동 분류 대비 빠른 분류 시간
- 창구 영업 응대 시간 감소 및 효율성 증대
- 오인 없는 문서 분류

➔ 문서 분류 확대 활용

- 추가적인 문서에 대한 분류 학습 및 적용 가능
- Email, 파일서버 등 디지털 문서 분류

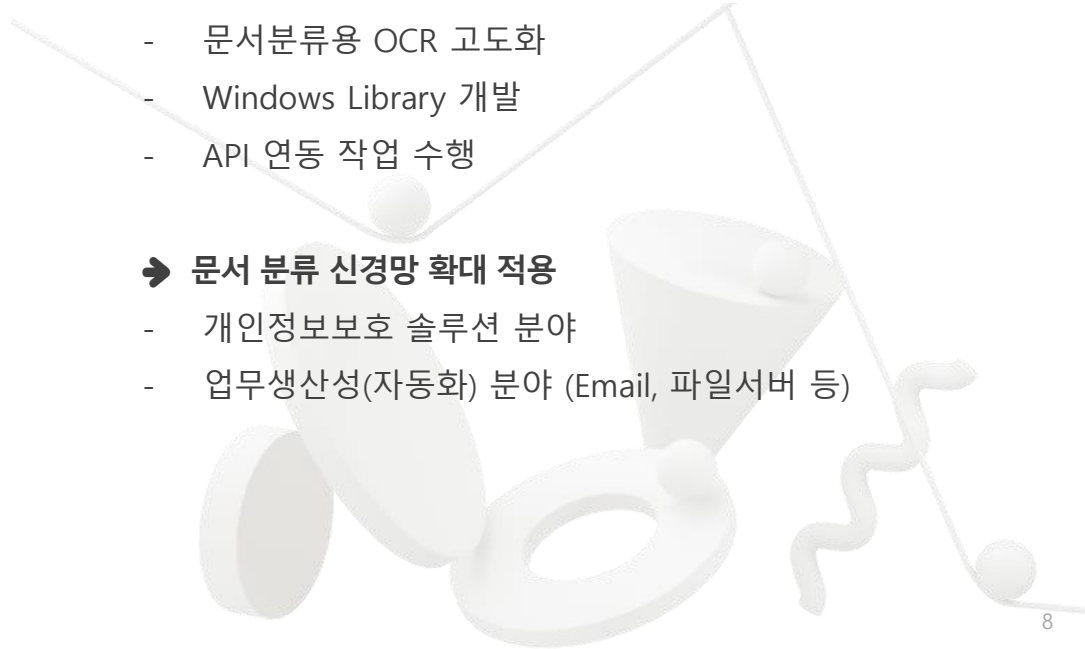
COMTRUE[®]
Technologies

➔ 인공지능 문서 분류 신경망 개발

- 자체 인공지능 문서 분류 신경망 개발
- 문서분류용 OCR 고도화
- Windows Library 개발
- API 연동 작업 수행

➔ 문서 분류 신경망 확대 적용

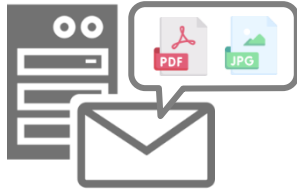
- 개인정보보호 솔루션 분야
- 업무생산성(자동화) 분야 (Email, 파일서버 등)



기대효과

OCR 활용 기술 (이메일, RPA)

- ✓ 이메일 첨부파일 내 이미지 키워드 인식 및 취합
- ✓ 신분증 OCR, 주요 문서 OCR 및 RPA 연동

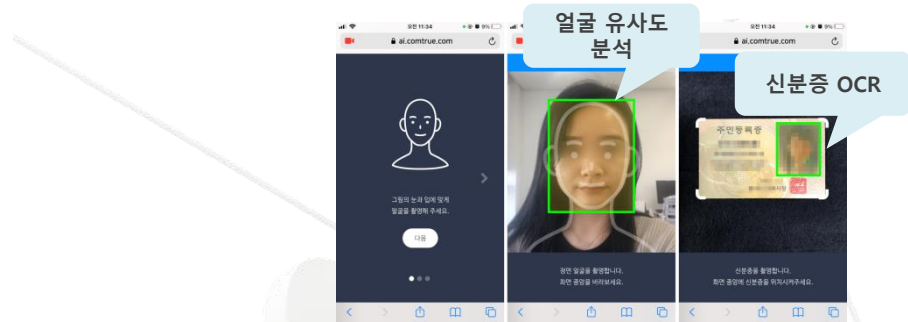


이메일 이미지 첨부파일 내



금융사 eKYC(비대면 본인확인, 신분증 OCR)

- ✓ 얼굴인증(유사도분석) 및 신분증 OCR, 진위확인을 제공
- ✓ 금융, 핀테크 기업의 비대면 본인확인 수행 지원



기대효과

문서 OCR

- ✓ 문서 내 주요 항목에 대한 OCR 제공
- ✓ [지원] 신분증4종, 신용카드, 개인 공문서 6종, 기업문서 3종

구분	항목명 (1차)	구분	항목명 (2차)
개인	주민등록증	개인	의료비영수증 OCR
개인	운전면허증	기업	통신판매업신고증 OCR
개인	여권	기업	영업신고증 OCR
개인	외국인등록증	기업	법인인감증명서 OCR
공통	신용카드	기업	사업자등록 증명원
개인	가족관계증명서 OCR	기업	부가세과세표준 증명원
개인	기본증명서 OCR	기업	지방세완납증명
개인	주민등록표 초본 OCR	기업	국세완납 증명원
개인	주민등록표 등본 OCR	기업	4대보험 가입자명부
개인	개인인감증명서 OCR	기업	표준재무제표 증명원
개인	본인서명사실확인서		
기업	사업자등록증 OCR		
기업	법인인감증명서 OCR		
기업	법인등기부등본 OCR		

이미지 개인정보보호 (문서 분류)

- ✓ 주민등록번호 등 이미지 내 주요 개인정보 스캔 (OCR 이미지 스캔)
- ✓ [+문서 분류 기능] 이름, 생년월일, 주소 등 기존 개인정보 필터링이 어려운 항목에 대해서도 사전 분류를 통해 개인정보보호 수행

 2021 개인정보 보호활용 기술 챌린지 개인정보위원장상 수상



이름, 주소, 생년월일이 포함된
개인정보 공문서 검출

감사합니다.

COMTRUE[®]
Technologies

(주)컴트루테크놀로지

- 회 사 명 : (주)컴트루테크놀로지
- 홈페이지 : <http://www.comtrue.com> | 인공지능 서비스 브랜드 : <https://cometrue.ai>
- 주 소 : 서울본사 (03920) 서울시 마포구 성암로 330(상암동 1580번지, DMC철단산업센터) 4층
세종지사 (30026) 세종특별자치시 조치원읍 터미널안길 60(상리 96번지, 세종고용복지센터) 6층
- 연 락 처 : 서울본사 02-396-7005(제휴 304 | 제품문의 505) ask@comtrue.com