## 포지큐브(AI-2) - 문서 OCR 기술 개발 채용 평가

- 1. 테스트 내용 : 한글/숫자/영문으로 이루어진 Character 인식 학습 모델 개발
  - 제공되는 데이터 셋을 활용하여 딥러닝/머신러닝 등의 모델을 설계하고 학습을 진행(4번 참고)
  - Validatation셋 정확도 측정(4번 참고)
  - 학습된 모델이 동작하는 환경(Weights 파일 포함) 제공(3번 참고)
  - 동작 방법 정확도 결과가 기재된 문서 제공
  - 소스코드 제공, 학습 결과 Weights 파일 제공
  - 웹사이트 참고 가능(참고한 사이트가 있으면 반드시 별도 기재)
- 2. 테스트 결과 제출 기간 : 채용 평가 메일 수신일로부터 7일 이내
  - 기간 내 제출 어려운 사유 발생 시 사유와 함께 제출이 가능한 기간 기재 후 담당자에게 연락 (7번 참고, 내부 확인 후 피드백 예정)
- 3. 동작 환경 :: Binary 또는 Anaconda 가상환경 또는 Docker 이미지
  - 개발 및 학습 환경 : 무관(가능하면 Python/Pytorch 환경 우선)
  - 테스트 환경: Ubuntu 16.04 또는 Windows/macOS 최신 버전에서 동작하는 환경 제공
  - 동작 방식: Validataion 셋과 동일한 형태의 Test 셋 폴더 지정 시 정확도/속도 출력

(속도 : 모델 로딩부터 Prediction 전체 시간, ms 단위, 정확도:4번 참고)

## 4. 학습 상세 :

- 딥러닝 프레임워크(pytorch, tensorflow, keras, darknet) 또는 머신러닝 기법(opencv, dlib 활용)을 사용
- 학습을 위한 모델 설계 : 가능한 독자 설계(알려진 모델 네트워크를 활용 가능하지만, 사용 이유 명시)
- 학습 방식: Pre-train된 Weights 사용 제한, Full-train으로 학습
- 제공되는 Train 셋을 활용하여 학습 진행(Data Augmentation 자유, 학습 데이터에 Validation 셋 포함 안됨)
- Validataion 셋을 이용하여 정확도 측정 후 결과 제공(정확도 : 정답 데이터 개수/전체 데이터 개수)

## 5. 학습 데이터 상세

- Train("train.zip"): 이미지 24150 개 (64x64 흰색배경)
- Validation("validation.zip"): 이미지 24150 개 (32x32 배경노이즈, 랜덤 마진, -2 ~ +2도 회전)
- Test("test.zip", 제공되지 않음): 이미지 24150 개 (32x32 배경노이즈, 랜덤 마진, -2 ~ +2도 회전)
- Ground Truth : 각 폴더 내에 "qt.txt" 참고
- Font("font.zip": 해당 데이터 생성에 사용된 Font 모음
- 데이터 셋에 사용된 문자 집합 : 마지막 페이지 "- 학습 데이터 문자 집합 -" 참고
- Data Augmentation : 필요에 따라 수행
  - 1) 제공되는 Train 셋을 이용하여 이미지 전처리
  - 2) 제공되는 폰트를 활용 추가 데이터 생성
  - \* 참고 사이트 : https://github.com/Belval/TextRecognitionDataGenerator.git

## 6. 평가 기준

- 채용 평가 진행 여부 및 수행 과정 평가(중요): 인터뷰 시 상세 내용 질문 후 평가
- 정확도 및 속도 평가
  - 1) 정확도: Validation, Test 셋 기준 낮은 결과
  - 2) 속도: Test 셋 10회 수행 평균
- 모델 설계 : 인터뷰 시 모델 설계(사용)에 대한 이유 질문 후 평가
- Data Augmetation :
  - 1) 진행하는 경우 방법 이유에 대해 인터뷰 시 질문 후 평가
  - 2) 진행하지 않는 경우 이유에 대해 인터뷰 시 질문 후 평가
- 코드 리뷰 : 인터뷰 시 추가 질문 후 평가
- \* 평가 과제를 완료하지 못하더라도 중간 결과물을 제공한 경우 해당 내용으로 평가(중요)
- 7. 문의 사항 및 결과물 제출 : 아래 메일을 통해 문의 또는 제출
  - 평가/일정 문의 및 결과물 제출 : niki@posicube.com
  - 기타 채용 절차 문의 : recruit@posicube.com

- 학습 데이터 문자 집합 -

"ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789가각간갇갈갉갊감갑값갓갔강갖갗같갚갛개객갠갤 갬갭갯갰갱갸갹갼캴걋걍걔걘걜거걱건건걸걺검겁것겄겅컺겉겊겋게겐곌겜겝겟겠겡겨격켞견겯**결**겸겹겻겼경곁계곈곌곕곗고곡곤곧 골꼶곬곯곰곱곳공곶과곽관괄괆괌괍괏광괘괜괠괩괬괭괴괵괸괼굄굅굇굉교굔굘굡굣구국군굳굴굵굶굻굼굽굿궁궂궈궉권궐궜궝궤궷 귀母권궐궘귑귓규균귤그극근귿글긁금급긋긍긔기긱긴긷길긺김깁깃깅깆깊까깍깎깐깔깖깜깝깟깠깡꺝깨깩깬깰깸깹깻깼깽꺄꺅꺌꺼 꺽꺾껀껄껌껍껏껐껑께껙껜껨껫껭껴껸껼꼇꼈꼍꼐꼬꼭꼰꼲꼴꼼꼽꼿꽁꽂꽃꽈꽉꽐꽜꽝꽤꽥꽹꾀꾄꾈꾐꾑꾕꾜꾸꾹꾼꿀꿇꿈꿉꿋꿍꿎 꿔꿜꿨꿩꿰꿱꿴꿸뀀뀁뀄뀌뀐뀔뀜뀝뀨끄끅꼰끊끌뀲뀷끔끕꼿끙끝끼낔꺈낄낌꺕낏낑낟낰낚낟낟날낡낢남납낫났낭닺밫낱놯냬냌낸낼 냄냅냇냈냉냐냑냔냘냠냥너넉넋넌널넒턻넘넙넛넜넝녛네넥넨넬넴넵넷넸넹녀녁년녈뎜녑녔녕녘뎨녠노녹논놀놂놈놉놋농높놓놔놘놜 놨되뇐뇔뇜뇝됫됴뇩뇬뇰뇹뇻뇽누눅눈눋눌눔눕눗눙눠눴눼뉘뒨뉠뉨뉩뉴뉵뉼늄늅늉느늑는늘듥늚늠늡늣틍흦늪늬늰늴니닉닌닐닒딤 닙닛닝닢다닥닦단닫달닭닮닯딿담답닷닸당닺닻돻대댁댄댈댐댑댓댔댕댜더덕덖던덛덜덞덟덤덥덧덩덫덮뎨덱덴델뎀뎁뎃뎄뎅뎌뎐뎔 뎠뎡뎨뎬도독돈돌돎돐돔돕돗동돛돝돠돤돨돼됐되된될됨됩됫됬됴두둑둔둘둠둡둣둥둬뒀뒈뒝뒤뒨뒬뒵뒷뒹듀듄듈듐듕드득든들들 듦듞듭틋틍듸다탁단뒫뒬뒴뒵팃팄팅툊따딱딴똴땀땁땃땄뙁뙇때땍떈뙐땜땝땟땠땡떠떡떤뛀떪떫떰떱떳떴떵뛓뗴떽뗀뛜뗌뗍뗏뗐뗑뗘 뗬도똑똔똘똥똬똴뙈뙤뙨뚜뚝뚠뚤뚫뚬뚱뛔띢뛴뛀뜀뛉뜅뜨띀뜬뜯뜰뜲뜹뜻띄띈띌띔띕띠띤띨띰띱띳띵라락란랄람랍랏뢌랑랓랖랗래 랙래랠램랩랫랬랭랴럌랸럇량러럭런럴럼럽쥣렀렁렇레렉렌렐렘렙렛렝려력련렬렴렵렷렸령례롄뤱롓로록론롤록롭롳롱롸뢌뢍뢨뢰뢴 뢸뢲룁룃룅료룐룔룝룟룡루룩룬룱룪룮룻룽뤄뤘뤠릮뤽륀륄륌륏륑류륰륜률륝륩륫륞르륵른릌름릅릊릉릊릍릁리릭린릴림립릿링마막 만많말말맑맒맘맙맛뫙맞땉맣매맥매맬맴맵맸맸맹맺먀먁먈먕더먹면멀멂멈멉멋멍멎멓메덱뭰멜멤멥멧멨멩며멱면멸몃몄명몇몌모목 몫몬몰몲몸몹못몽뫄뫈뫘뫙뫼묀묄묍묏묑묘묜묠묩묫무묵묶문묻물묽묾뭄뭅뭇뭉뭍뭏뭐뭔뭘뭡뭣뭬뮈뮌뮐뮤뮨뮬뮴뮷므믄믈믐믓미믹 <u>민</u>믿밀밂밈밉밋밌밍믳밑바박봒밗반봗봘봙봚봛봠밥밧봥봩배백뱬봴뱀뱁뱃뱄뵁뵅뱌뱍뱐뱝버벅번벋벌벎범법벗벙볒<u>베</u>벸뻰벧뷀뛤벱 벳벴벵벼벽변별볍볏볐병볕볘볜보복볶본볼봄봅봇봉봐봔봤봬뵀뵈뵉뵌뵐뵘뵙뵤뵨부북분불붉붊붐붑붓붕붍붚붜붤붰붸뷔뷕뷘뷜뷩 뷰뷴뷸븀븃븅브븍븐블븜븝븟비빅빈븰빎빔빕빗빙븾빛뺘빡뺜뽤뽦빰빱빳뽰뽱뽷뺴빽뺸뺼빽뻅뺏뺐뺑뺘뺙뺨뻐뻑뻔삗뻘뻠뻣뻤뻥뻬뼹 뼈뼉뼘뼙뼛뼜뼝뽀뽁뽄뽈뽐뽑뽛뾔뾰뿅뿌뿍뿐뿔뿜뿟뿡쀼쁑쁘쁜쁠쁨쁩뻬삑삔삘삠삡삣삥사삭삯산삳살삵삶삼삽삿샀상샅새색샌샐샘 샙샛샜생샤샥샨챨샦샵샷샹섀섄섈섂섕서석섞섟선섣설섦섧섬섭섯젔성섶세섹센셀셈솁셋셌셍셔셕션셜셤셥셧셨셩졔셴셸솅소속솎손 솔솖솜솝솟송솥솨솩솬솰솽쇄쇈쇌쇔쇗쇘쇠쇤쇨쇰쇱쇳쇼쇽숀숄숌숍숏숑수숙순숟술숨숩숫숭숯숱숲숴쉈쉐쉑쉔쉘쉠쉥쉬쉭쉰쉴쉼쉽 쉿슁슈슉슐슘슛슝스슥슨슬슭슦습츳승시식신칟실싫심십싯징싶싸싹싻싼쌀쌈쌉쌌쌍짷쌔쌕쌘쌜쌤쌥쌨쌩썅써쎀쎤썰썲쎰썹썼쎵쎄쎈 쎌쏀쪼쏰쫀쫃쏠쏢쏨쏩쏭쫘쐌쫜쫬쐐쐤쐬쐰쐴쬠쐽쬬쑤쑥쭌쑬쑴쑵쭝쒀쒔쒜쒸쒼쓩쯔쓱쓷쓸쓺쓿씀씁씌씐씔씜쌰앀쌴씰씸씹찟씽아앜 안앉않알앍앎앓암압앗았앙앝앞애액앤앨앰앱앳앴앵야약얀얄얇얌얍얏양얕얗얘얜얠얩어억언얹얻얼얽얾엄업없엇었엉엊엌엎에엑엔 엨엒엡엣엥여역엮연열엶엷염엽엾엿였영옅옆옇예옌옐옘옙옛옜오옥온올옭옮옰옳옴옵옷옹옻와왁완왈왐왑왓왔왕왜왝왠왬왯왱외왹 왼욀욈욉욋욍요욕욘욜욤욥욧용우욱운울욹욺움웁웃웅워웍원월웜웝웠웡웨웩웬웰웸웹웽위윅윈윌윔윕윗윙유육윤율윰윱윳융윷으윽 은을읊음읍읏응읒웇읔읕읖읗의읜읠읨읫이익인일읽읾잃임입잇있잉잊잎자작잔잖잗잘잚잠잡잣잤장잦재잭잰잴잼잽잿쟀쨍쟈쟉쟌쟎 쟐쟘쟝쟤쟨쟬저적전절젊점접젓정젖제젝젠젤젬쩹젯젱져젼졀졈졉졌졍졔조족존졸졺졲좁좃종좆좇좋좌좍좔좝좟좡좨좼좽죄죈죌죔죕 죗죙죠죡죤죵주죽준줔줅줆죾줍줏중줘줬줴쥐쥑쥔쥠쥠쥓쥬쥰쥴쥼즈즉즏즡즊즙즛증지직진짇질짊짐집짓징짗짙짚짜짝짠짢짤짧짬 짭짯짰짱째짹짼쨀쨈쨉쨋쨌쨍쨔쨘쨩쩌쩍쩐쩔쩜쩝쩟쩠쩡쩨쩽쪄쪘쪼쪽쫀쫄쪾쫍쫏쫑쫓쫘쫰쫠쫬퐤쬈쬐쬔쬘쬠쬡쭁쭈쭉쭌쭐쭊쭙쭝쭤 쭸쭹쮜쮸쯔쯤쯧쯩찌찍찐찔찜찝찡찣찧차착찬찮찰참찹찻찼창찿채챜첀첄첌챕챗챘챙챠챥챦챨챰챵처첰쳔첰쳠첩첫첬청체첵쳰쳴쳄쳅 쳇쳉쳐쳔쳤쳬쳰촁초촉촌촐촊촙촞총촤촨촬촹최쵠쵤쵬쵭쵯쵱쵸춈추춖춘출춖츕춧충춰췄췌췐취췬췰췸췸췻췽츄츈츌츍츙츠츸츤츨츢 츱츳층치칙친칟칠칡침칩칫칭카칵칸칼캄캅캇캉캐캑캔캘캠캡캣캤캥캬캭컁커컥컨컬컬컴컵컷컸컹케켁켄켈켐켑켓켕켜켡켤켬켭켯컜 켱켸고콕콘콜콤콥콧콩콰콱콴콸쾀쾅쾌쾡쾨쾰교쿠쿡쿤쿨쿰쿱쿳쿵쿼퀀퀄퀑퀘퀭퀴퀵퀸퀼큄큅큇큉큐큔큘큠크큭큰클큼큽킁키킥킨킬 <u>킴킵킷킹타탁탄탈탉탐탑탓탔탕태택턘턜탬탭탯탰탱탸턍더텈턴털턺텀텁텃텄텅테텍텐텔템텝텟텡텨텬텼뎨톈토톡톡톨톩톱톳톻톺톼</u> 퇀퇘퇴퇸툇툉툐투툭툰툴툼툽툿퉁퉈퉜퉤튀튁튄튈튐튑튕튜튠튤튬튱트특튼틀틂틈틉틋틔틘틜틤틥티틱틴틸팀팁팃팅파팍팎판팔팖 팜팝팟팠팡팥패퍀퍤팰팸팹팻팼팽퍄퍅퍼퍽펀펄펌펍펏펐펑페펙펜펠펨펩펫펭펴편펼폄폅폈평폐폘폡폣포폭폰폴폮폽폿퐁퐈퐝푀푄표 푠푤푭푯平푹푼푿풀풂품풉풋풁풔풩퓌퓐퓔퓜퓟퓨퓨퓰퓸퓻퓽프픈플픔픕픗피픽핀필핌핍핏핑하학한할핥햠합햣햠핳해햌해햌햄햅햿 했행햐향허헉헌헐헒험헙헛헝혜헥헨헬헴헵헷헹혀혁현혈혐협혓혔형혜혠혤혭호혹혼홀홅홈홉홋홓홑화확환활홧황홰홱홴횃횅회획흰 횔횝횟횡됴횬횰룝횻후훅흤훜훑훆훗훙훠훠훨훰훵훼훽휀휄휑힊휙휘휠휨휩휫휭휴휵휸휼휶흇휺흐흨츤흖흗흘츩흠츱흣흥흩희흰흴흼 흽힁히힉힌힐힘힙힛힝힣"