DEVIEW
2019

나 대신 손글씨 써주는 AI 만들기

(성공적인 Side Project)

이바도
CLOVA AI OCR
NAVER

쓰인 글꼴: 나눔손글씨 대한민국 열사체

CONTENTS

DEVIEW 2019

OCR 팀에서 왜 손글씨 폰트를 만들고 있었을까?: Synthetic Data 와 OCR

한글 폰트 만들기의 어려움: 영문 폰트 만들기와 한글 폰트 만들기의 비교

손글씨 생성 Prototype
- 생성 pipeline
- 손글씨 생성 모델
- 손글씨 생성이 유독 어려운 이유
- Pretrain 이 중요해

모델 성능 개선: Pretrain 부터 잘 하자!
- Sota / 최신 기법 때려 넣기
- 퀄리티 향상 결과

정말 개선되었을까?: 사용자 피드백

한글날 공모전
- 서비스를 위한 추가 요청 사항과 개선 사항들

성능 및 사용성 개선: 할 수 있는 것은 다 하자!
- 전처리 뜯어 고치기
- 글자 수 줄이기, 쉬운 글자로 바꾸기
- 비슷한 폰트로 시작하기

후처리 관련 내용들과 서비스 파이프라인

한글날 공모전 후기

What’s Next?

OCR 팀에서

왜

손글씨 폰트를

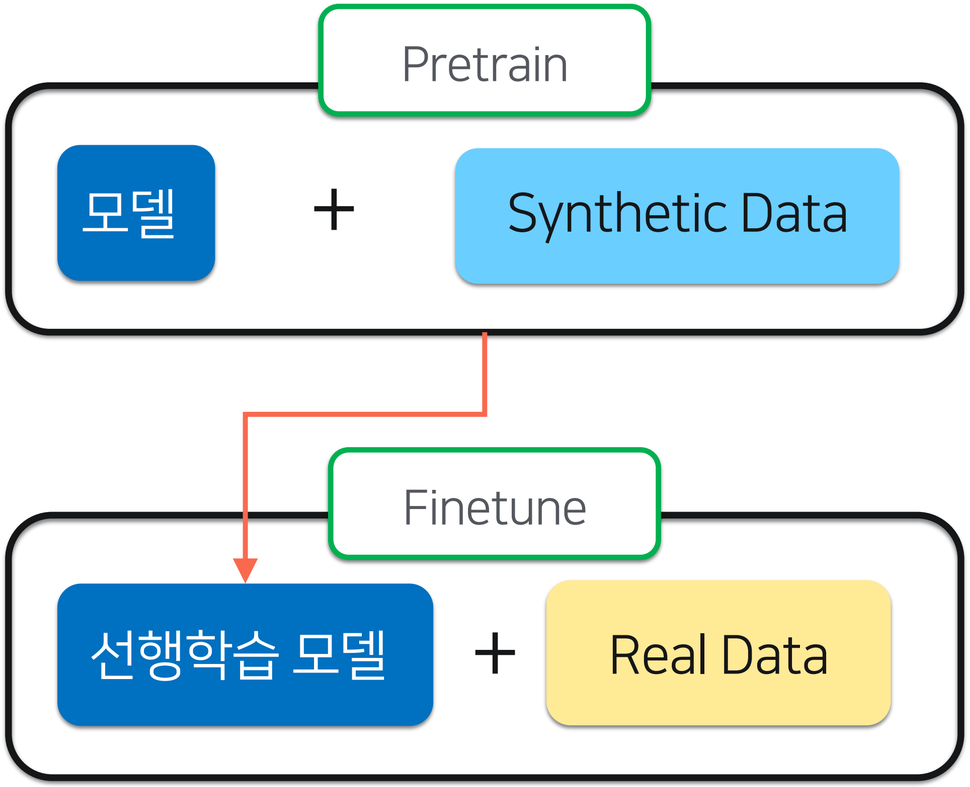
만들고 있었을까?

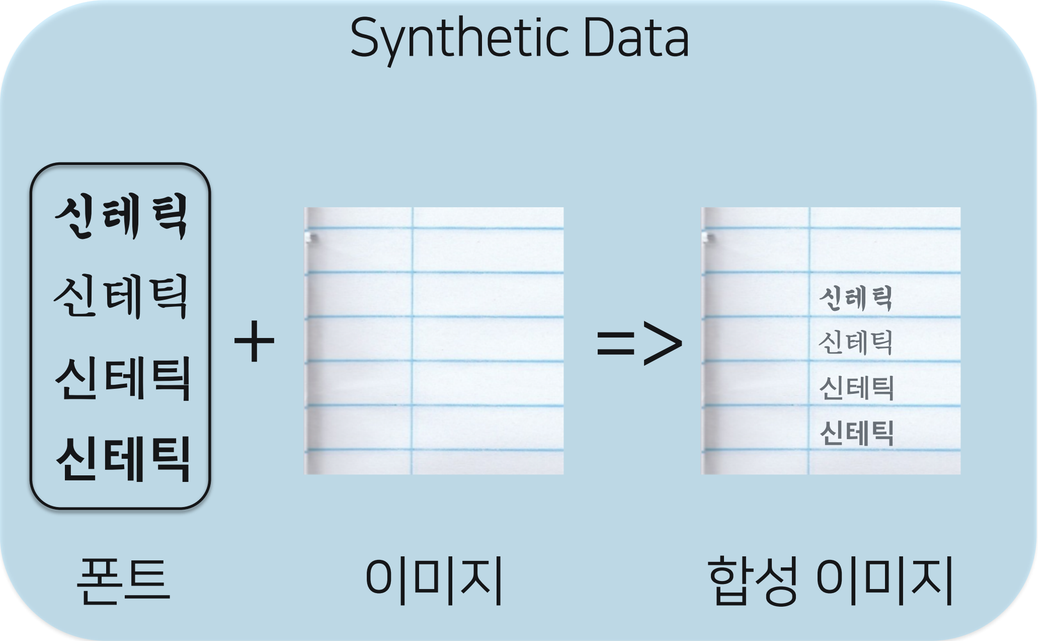
거기 그거 하는데 맞나요?

혹시... 하라는 일 안하고 딴짓 하는 거 아니야?

Synthetic Data 와 OCR

일반적인 활자체 OCR 학습

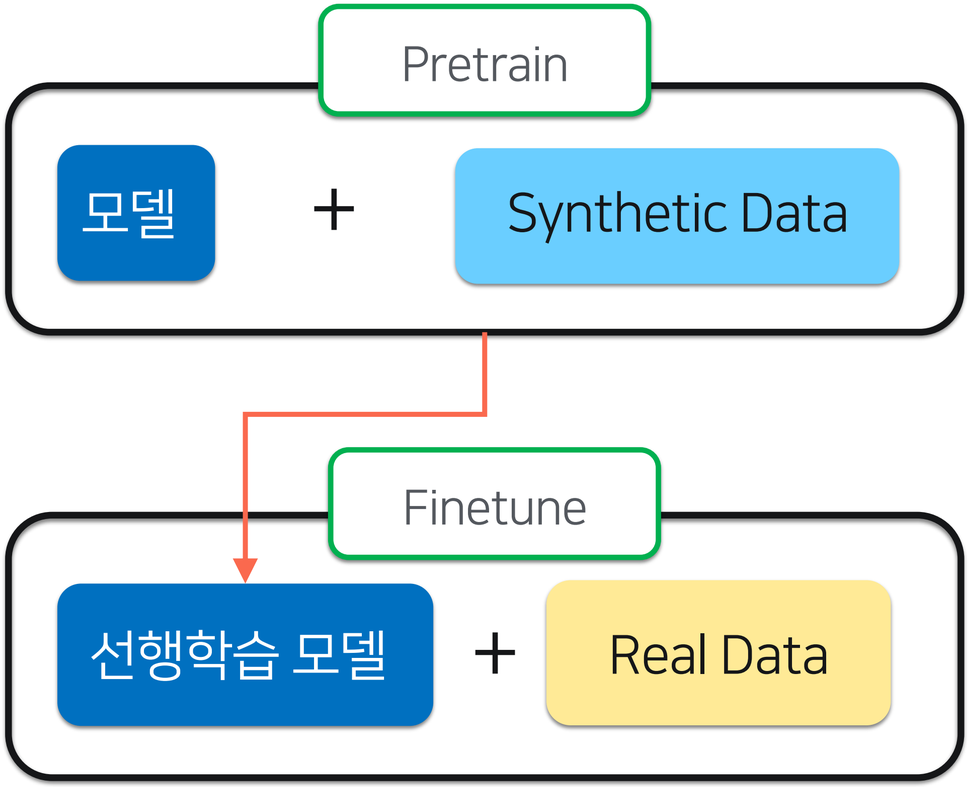


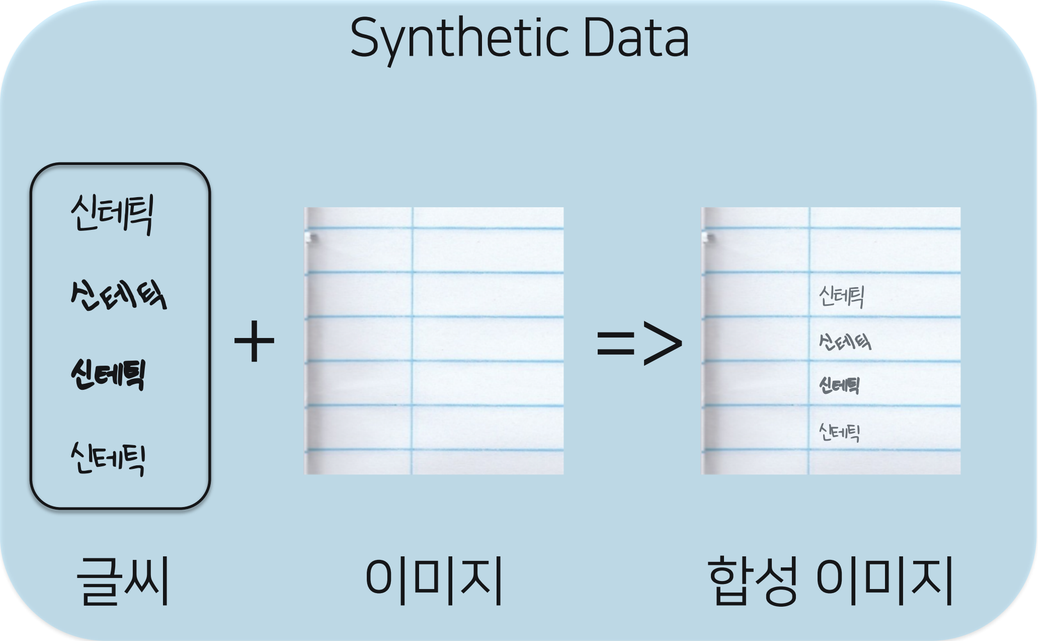


다량의 다양한 폰트 필요!

Synthetic Data 와 OCR

손글씨 OCR 학습





역시 다량의 다양한

손글씨 이미지 필요!

다량의 다양한

손글씨 폰트 찾습니다.

왜죠?

한글 한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

한글한글 아름답게
앞으로 쓰여질 우리의 이야기를 응원합니다.

아예 손글씨 폰트를 직접 만들어 보면 어떨까?

여러 사람한테 글씨 많이 많이 받아서

만들면 되는거 아냐?

얼마나 받으면 되는거지?

그러나,

한글 폰트 만들기의 어려움...

영어 폰트 만들기

0 1 2 3 4 5 6 7 8 9

! " # $ % & ' ( ) * + , - . / : ; < = > ? @ [₩] ^ _ ` { | } ~

ABCDEFGHIJKLMNOPQRSTUVWXYZ

abcdefghijklmnopqrstuvwxyz

숫자(10자) + 기호 (32자) + 알파벳 (26+26자) = 총 94자

한글 폰트 만들려면 몇 글자나 필요할까?

100 글자?

1000 글자?

10000 글자?

한글 폰트 만들기

영어와 기호 (94)

+ 낱 글자 (51)

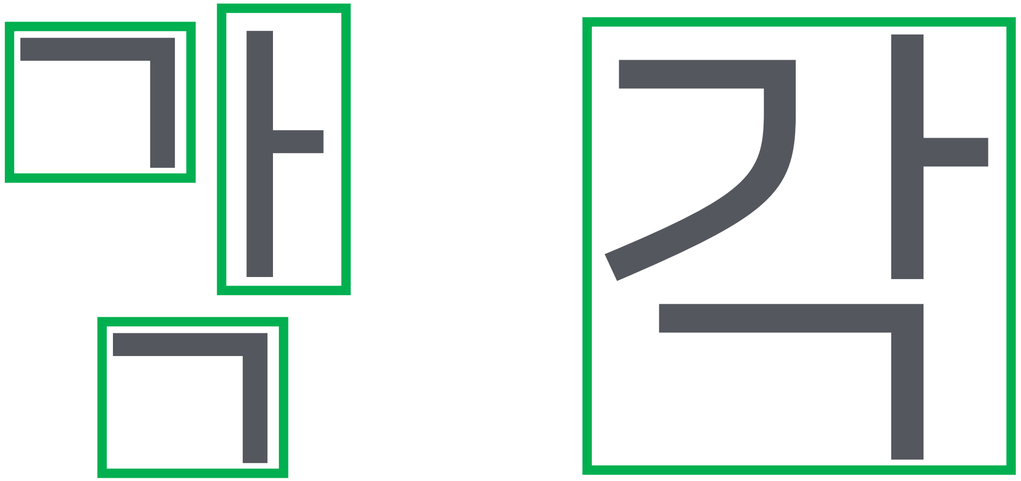
+ 초성 (19) x 중성 (21) x 종성 (27+1)

= 11,317 글자

200자 원고지 57장 분량

띄어쓰기 없이!

한글 폰트 만들기



조합형 완성형

한글 폰트 만들기

나눔스퀘어 Regular

한글 폰트 만들기

DEVIEW
2019

https://ko.wikipedia.org/wiki/파일:금속활자장.jpg

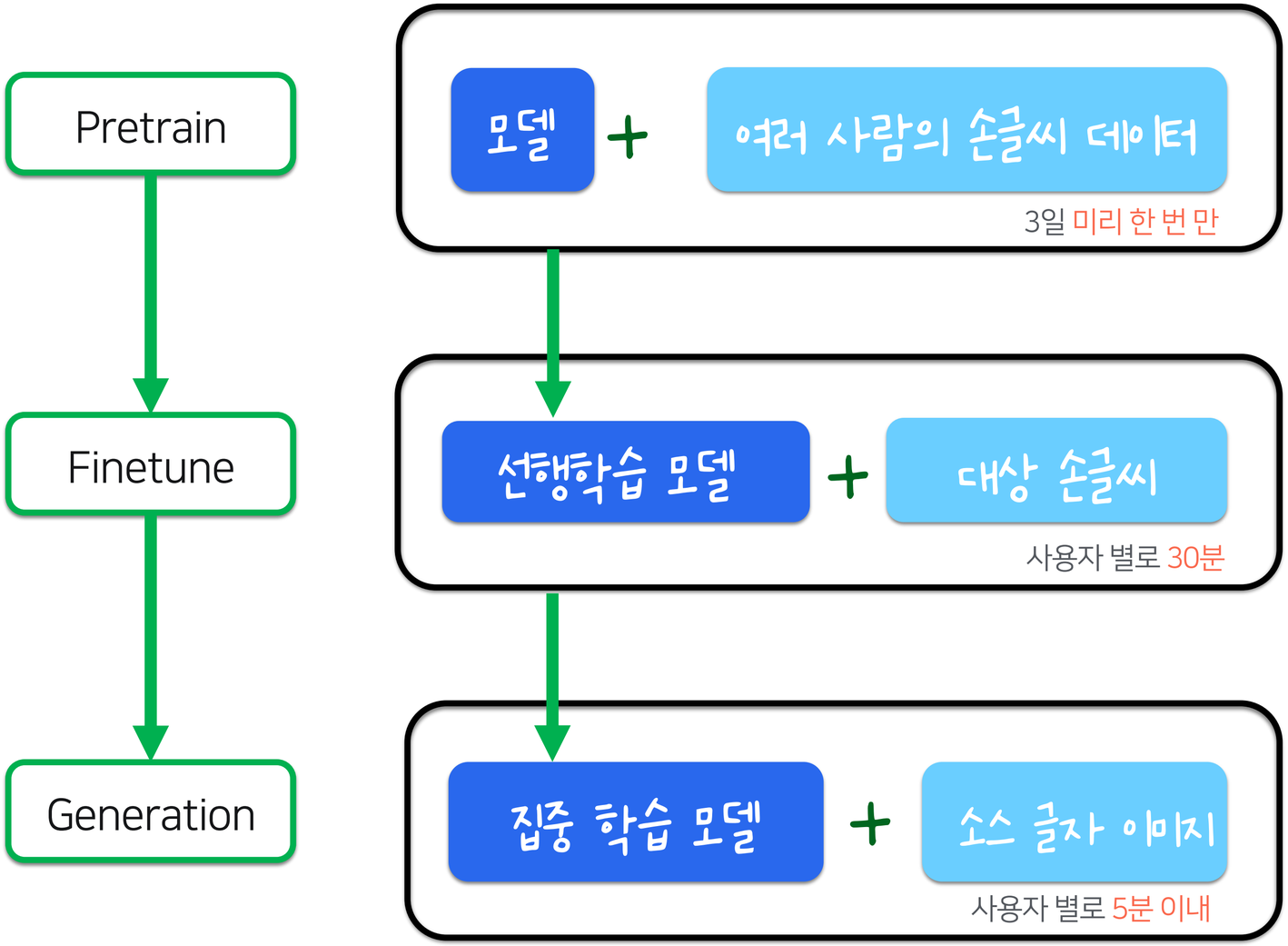
GAN, 많이 쓰던데?
아예 손글씨 폰트를 생성 모델로
생성해보면 어떨까?

그래서 한번 빠르게 만들어 봤습니다.

손글씨 생성

PROTOTYPE

생성 pipeline


Pretrain
Finetune
Generation
모델
+
여러 사람의 손글씨 데이터
3일 미리 한 번 만
선행학습 모델
+
대상 손글씨
사용자 별로 30분
집중 학습 모델
+
소스 글자 이미지
사용자 별로 5분 이내

손글씨 생성 모델 – 어떻게 생겼나?

DEVIEW 2019

Pix2pix

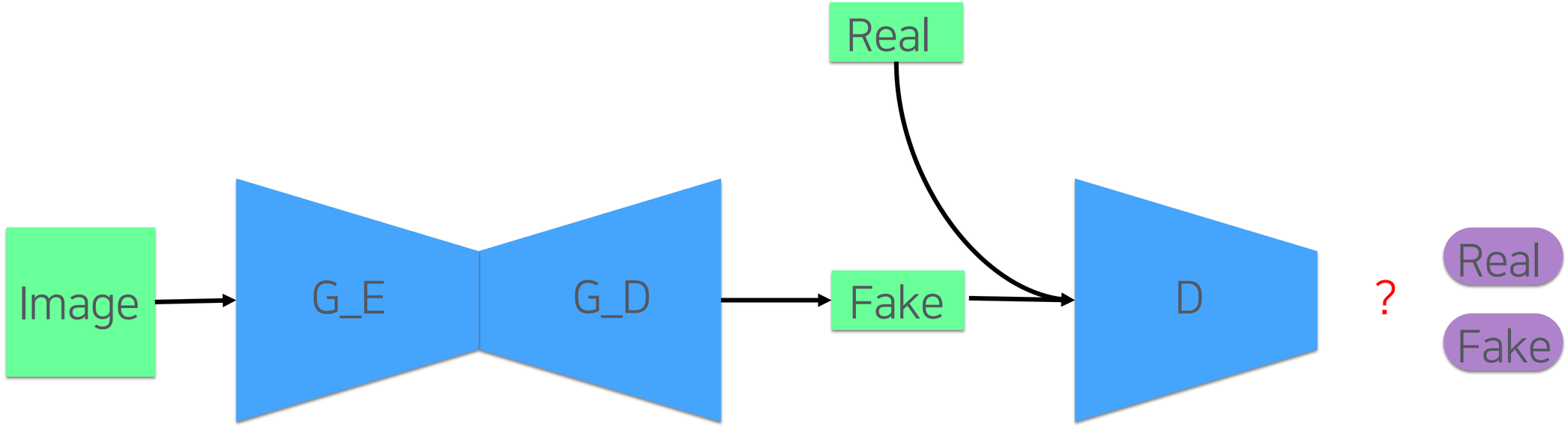
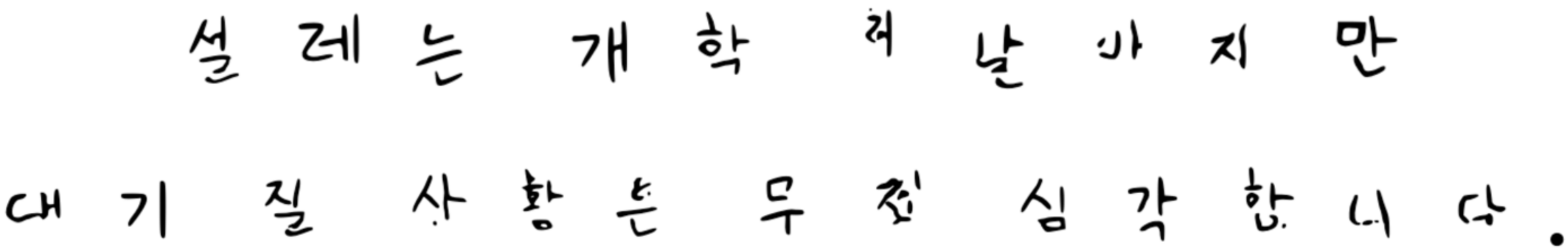


Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017

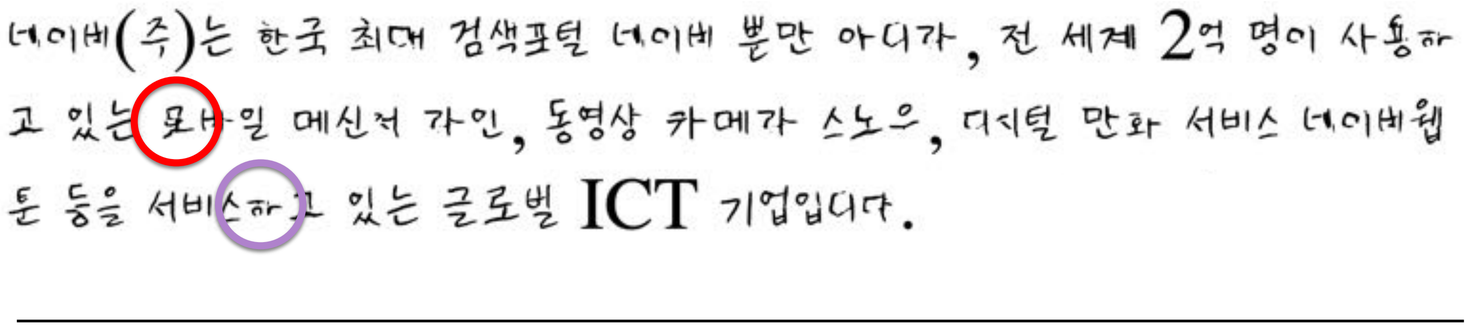
손글씨 생성 Proto Type 결과

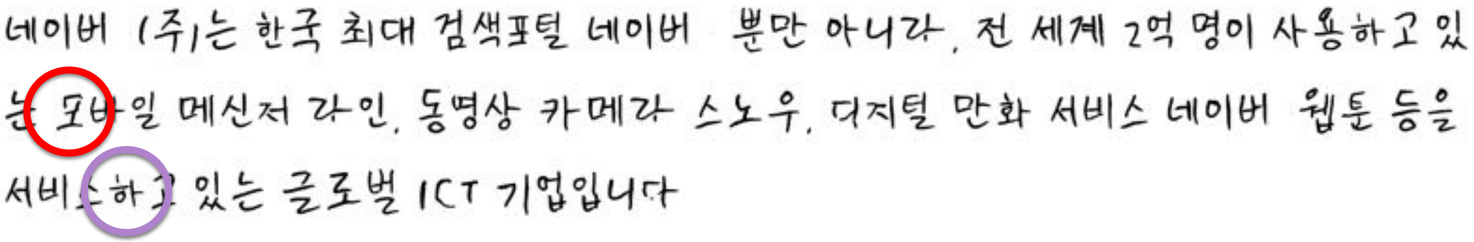


https://github.com/kaonashi-tyc/zi2zi
https://github.com/periannath/neural-fonts

손글씨 생성 Proto+ Type

Pretrain Data 교체





하지만 여전히 눈에 띄는 문제들

보 "ㅂ" 글자가 자라면 안되는 곳에서 자라나요

붊 "붊" 아무데나 점이 생겨요. ㅎ 닮은 ㄹ

쉴 히 "쉴", "히" 획이 사라져요

왜 손글씨 생성이 유독 어렵나요?

스타일이 섞여 있어요

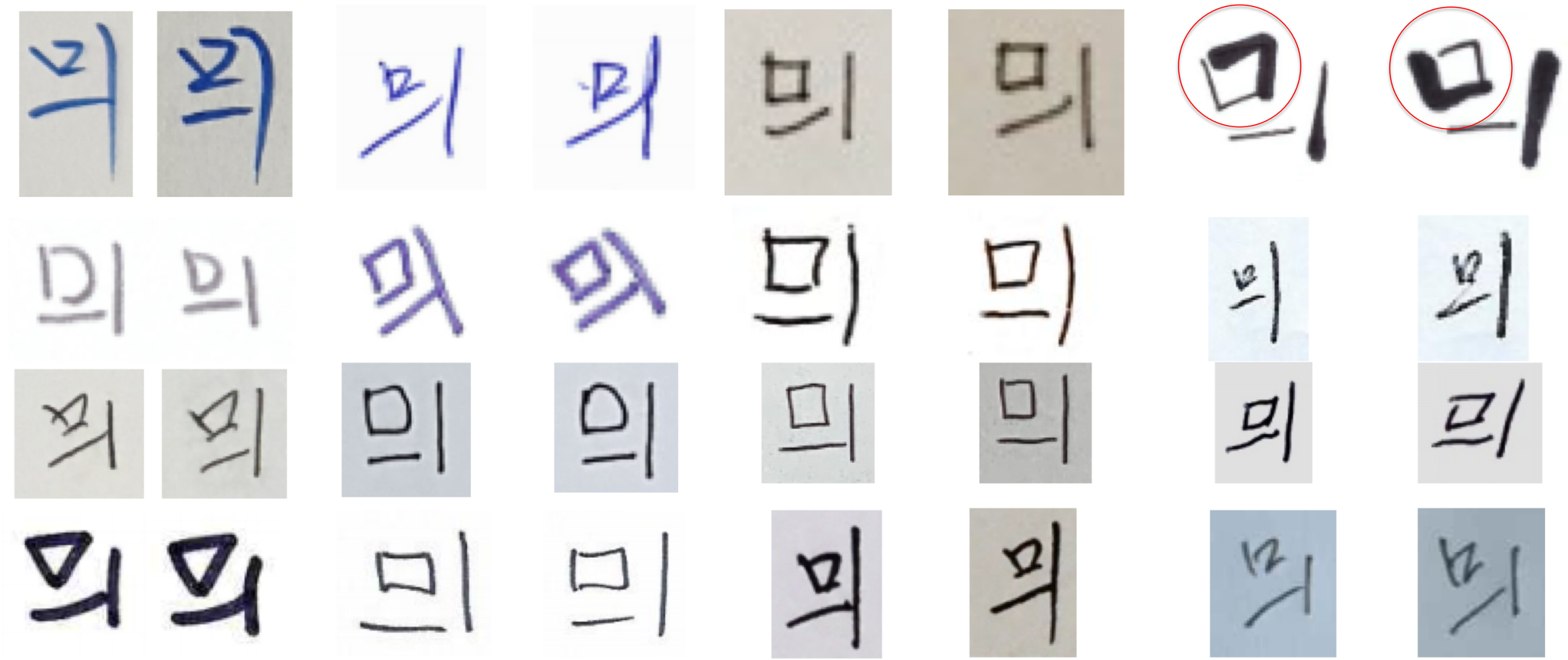
저는 고려 글꼴 입니다.

My English letters look quite different.

H H

똑같이 절대 못써요:

같은 글자는 두 번 다시 같게 쓰여지지 않는다...



쓰인 글꼴: 나눔손글씨 조석체

생성 모델이 불안정해요

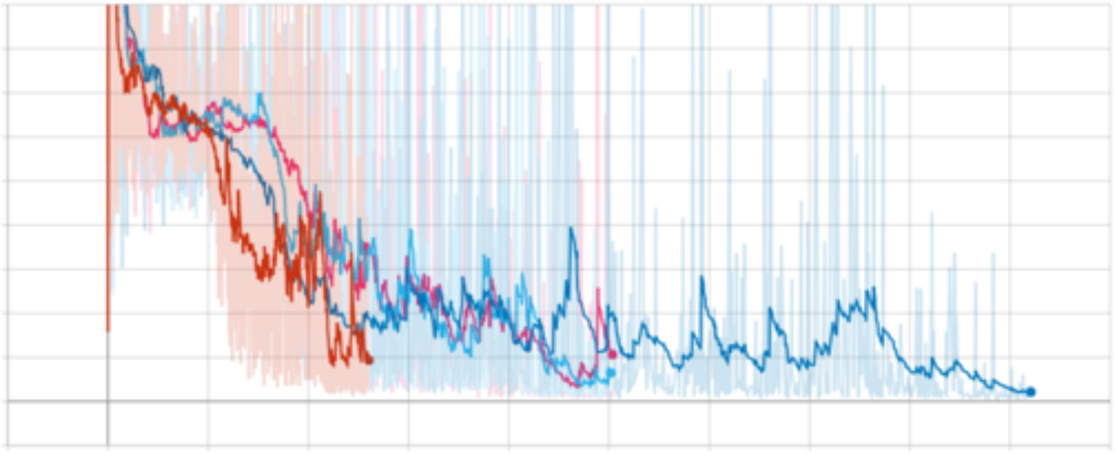
학습을 어디서 끊어야 할지

잘 모르겠어요.

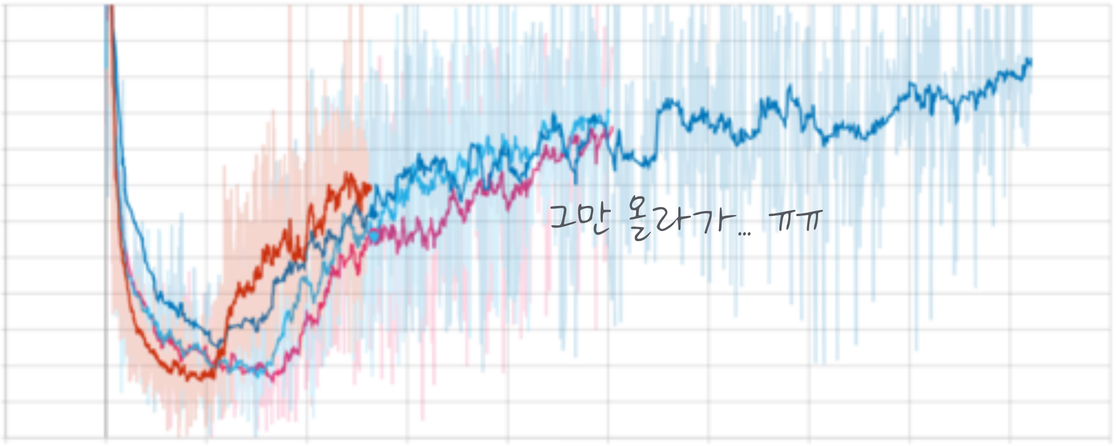
ART 에 가까움.

장인의 손맛? 감?

Discriminator loss

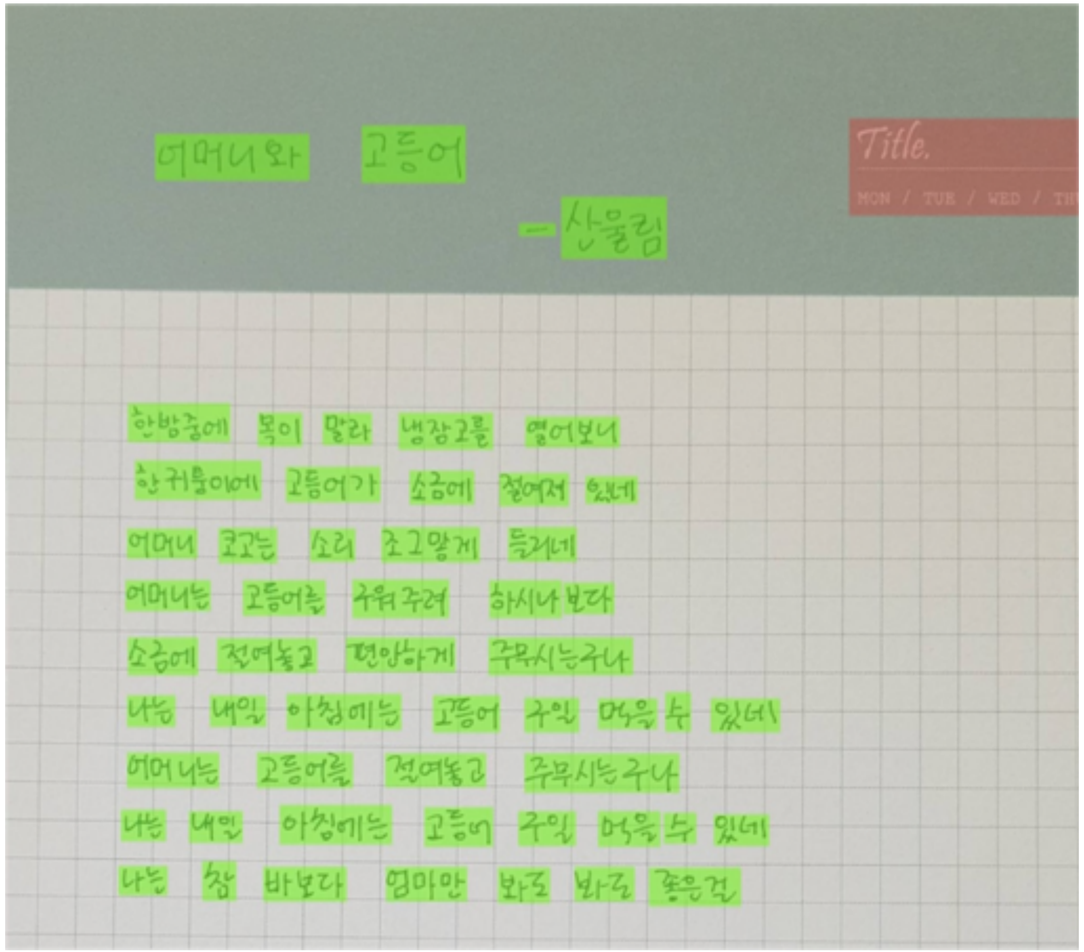


Generator loss



쓰인 글꼴: 나눔손글씨 기쁨밝음

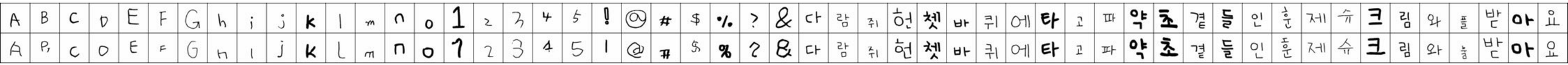
Labeling 된 손글씨 데이터가 잘 없어요

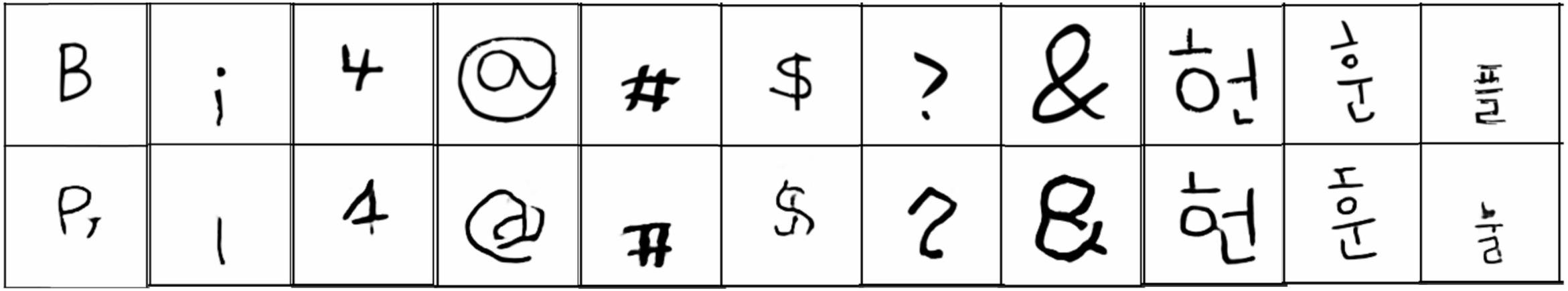


쓰인 글꼴: 나눔손글씨 가람연꽃

뭐부터 고쳐야 할까요?

Proto Type – Pretrain 초기 성능





모델 성능개선

Pretrain부터 잘 하자!

1차 성능개선: Pretrain 부터 잘 하자

안정적으로 (틀린 글자 없이)

스타일을 잘 반영(비슷하게) 하도록 학습해야 함

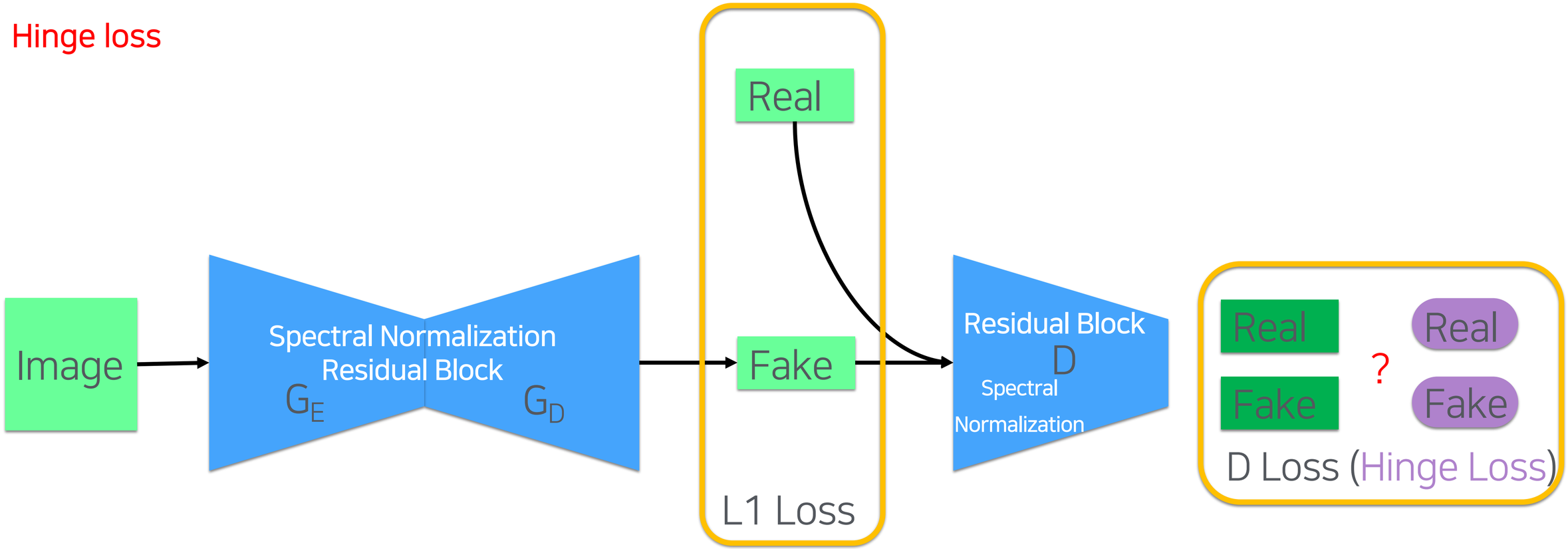
1차 성능개선: Pretrain 부터 잘 하자

DEVIEW 2019

Spectrally Normalized Gan (SNGAN)

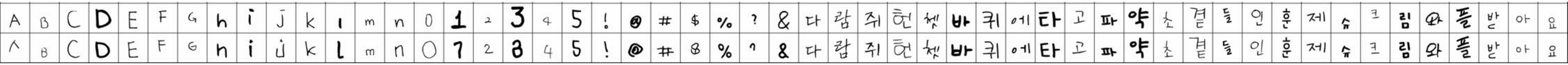
Residual block

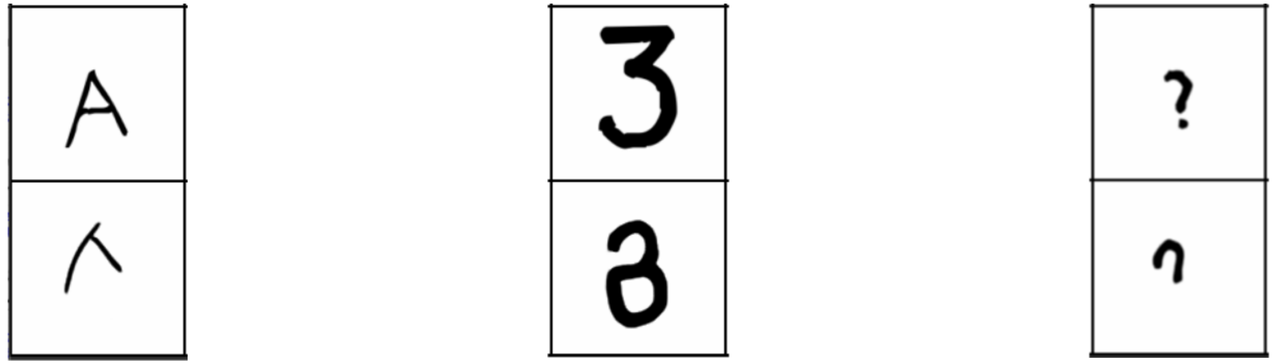
Hinge loss



pix2pix
+ Spectral Normalization
+ Residual block
+ Hinge loss

초기의 트레이닝 안정화 결과





SOTA / 최신 기법 때려 넣기

1차 성능개선: Pretrain 부터 잘 하자

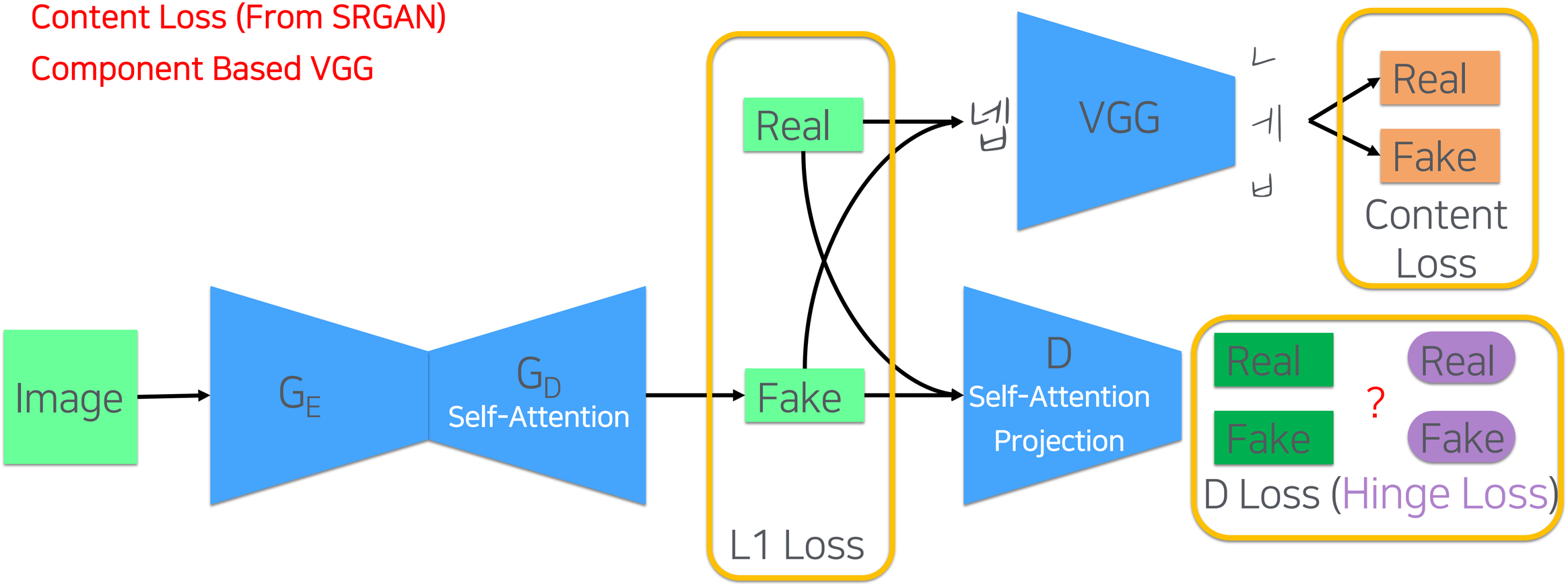
DEVIEW 2019

Self-Attention Gan (SAGAN)

Projection Discriminator

Content Loss (From SRGAN)

Component Based VGG



cGANs WITH PROJECTION DISCRIMINATOR, ICLR 2018

Self-Attention Generative Adversarial Networks, PMLR 2019

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, CVPR 2017

pix2pix
+ Spectral Normalization
+ Residual Block
+ Hinge Loss
+ Self Attention
+ Projection Discriminator
+ Perceptual Loss
+ Component Based VGG
+ R-Adam
+ instance norm
+ r1 penalty
+ orthogonal initialization
+ embed sharing
+ training variation
+ 그 외 다수의 실험들

퀄리티 향상 결과

정말 개선되었을까?

사람들도 그렇게 생각할까?

사용자 피드백

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나올실 때 황사용 마스크부터 챙기셔야겠는데요.
오늘은 강원 영동과 영남을 제외한 서쪽 지방을 중심으로 공기가
무척 탁하겠습니다.

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나오실 때 황사용 마스크부터 챙기셔야겠는데요.
오늘은 강원 영동과 영남을 제외한 서쪽 지방을 중심으로 공기가
무척 탁하겠습니다.

생성

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나올실 때 황사용 마스크부터 챙기셔야겠는데요.
오늘은 강원 영동과 영남을 제외한 서쪽 지방을 중심으로 공기가
무척 탁하겠습니다.

원본

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나오실 때 황사용 마스크부터 챙기셔야겠는데요.
오늘은 강원 영동과 영남을 제외한 서쪽 지방을 중심으로 공기가
무척 탁하겠습니다.

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나오실 때 황사용 마스크부터 챙기셔야겠는데요
오늘은 강원 영동과 영남을 제외한 서쪽 지방을 중심으로 공기가 무척
탁하겠습니다.

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나올실 때 황사용 마스크부터 챙기셔야겠는데요.
오늘은 강원 영동과 영남을 제외한 서쪽 비장을 중심으로 공기가 무척
탁하겠습니다.

원본

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나오실 때 황사용 마스크부터 챙기셔야겠는데요
오늘은 강원 영동과 영남을 제외한 서쪽 지방을 중심으로 공기가 무척
탁하겠습니다.

생성

설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.
나올실 때 황사용 마스크부터 챙기셔야겠는데요.
오늘은 강원 영동과 영남을 제외한 서쪽 비장을 중심으로 공기가 무척
탁하겠습니다.

Feedback

11명 중 10명(90.9%)의 사용자가
생성된 폰트에 대해 만족한다고 응답

성능이 좋아지니,

일이 더 늘어났다.

갑자기 분위기

한글날 공모전

한글날 공모전

공모전 기획 시작

2019년 7월 11일 목요일

안녕하세요 바도님, 글로바 기획 박원경이라고 합니다.
현재 손글씨 폰트 서비스 기획을 진행중인데요, 궁금한 점들이 몇가지 있어서 연락드립니다. :)

오후 5:18

읽음
오후 5:33

넵 안녕하세요! :)

한글날은 10월 9일...

세 달 남았네요...

공모전 오픈은 9월 4일...

두 달 남았군요... 어허허 어떻게든 되겠죠...

한글날 공모전 프롤로그 영상

한글날 공모전

그런데 말이죠...

스캐너 꼭 써야 하나요?

써야하는 글자가 많고 어려워요

혹시 추가적인 성능 개선의 여지는 없을까요?

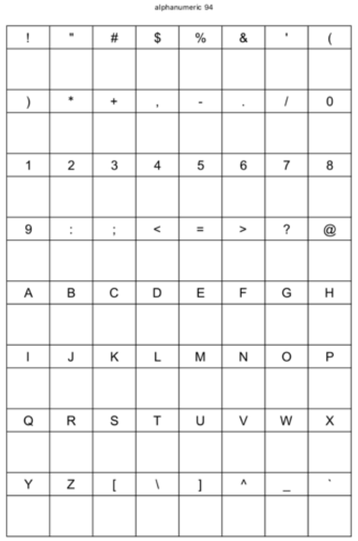
성능 및 사용성 개선
할 수 있는 것은 다 하자

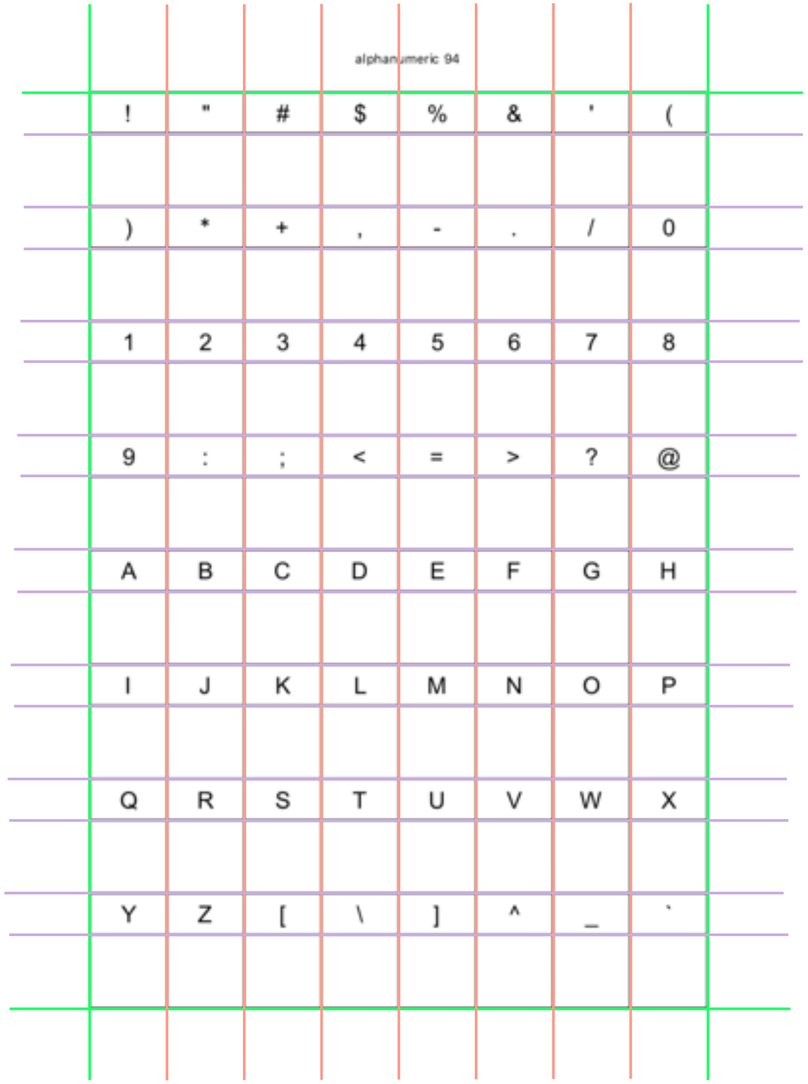
스캐너 꼭 써야 하나요?

전처리 뜯어 고치기

전처리 뜯어 고치기

초기 템플릿





템플릿 외곽선 검출

↓

등간격 분할

↓

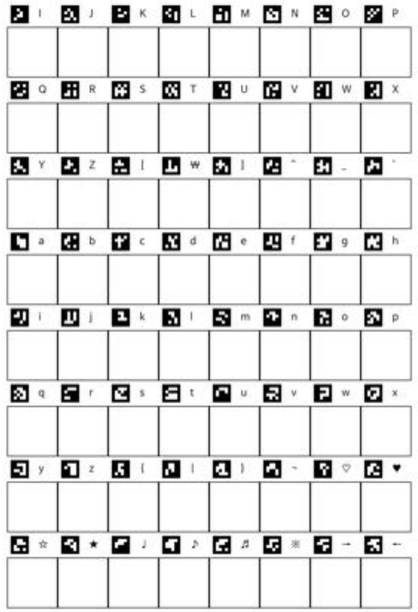
이미지 처리

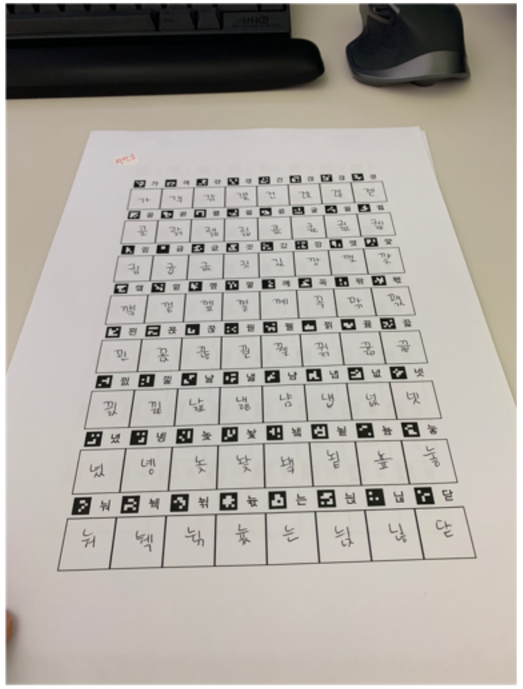
전처리 뜯어 고치기

초기 템플릿

전처리 뜯어 고치기

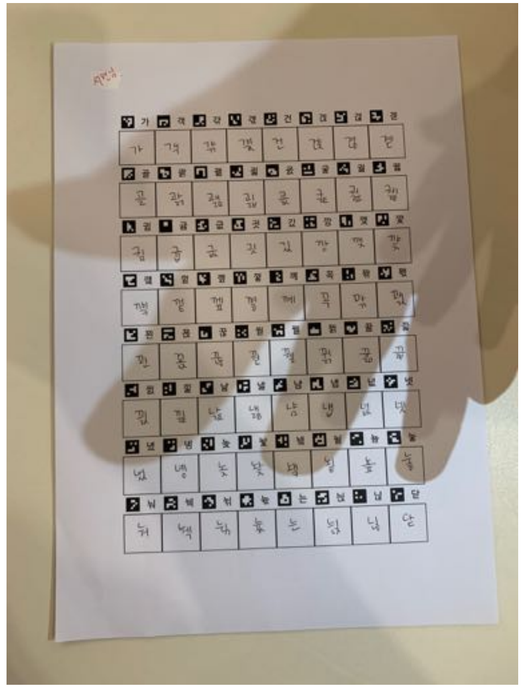
마커 템플릿 (OpenCV ArUco)





Perspective

Rotation



Shading

써야하는 글자가 많고 어려워요

글자 수 줄이기, 쉬운 글자로 바꾸기

글자 수 줄이기, 쉬운 글자로 바꾸기

꽤 뿲 싸 컑 풂 퉑
… 궷 랗 랖 래 랆 랖 …

글자 수 줄이기, 쉬운 글자로 바꾸기

원본 설레는 개학 첫날이지만 대기 질 상황은 무척 심각합니다.

400자 설레는 개학 첫날이지만, 대기질 상황은 무척 심각합니다.

200자 설레는 개학 첫날이지만, 대기질 상황은 무척 심각합니다.

100자 설레는 개학 첫날이지만, 대기질 상황은 무척 심각합니다.

50자 설레는 개학 첫날이지만, 대기질 상황은 무척 심각합니다.

10자 설레는 개학 첫날이지만, 대기질 상황은 무척 심각합니다.

글자 수 줄이기, 쉬운 글자로 바꾸기

DEVIEW 2019

초성 (19가지 type)

ㄱ ㄲ ㄴ ㄷ ㄸ ㄹ ㅁ ㅂ ㅃ ㅅ ㅆ ㅇ ㅈ ㅉ ㅊ ㅋ ㅌ ㅍ ㅎ

중성 (4가지 type)

ㅏ ㅐ ㅑ ㅒ ㅓ ㅔ ㅕ ㅖ ㅣ

ㅗ ㅛ ㅜ ㅠ

ㅘ ㅙ ㅚ ㅝ ㅞ ㅟ ㅢ

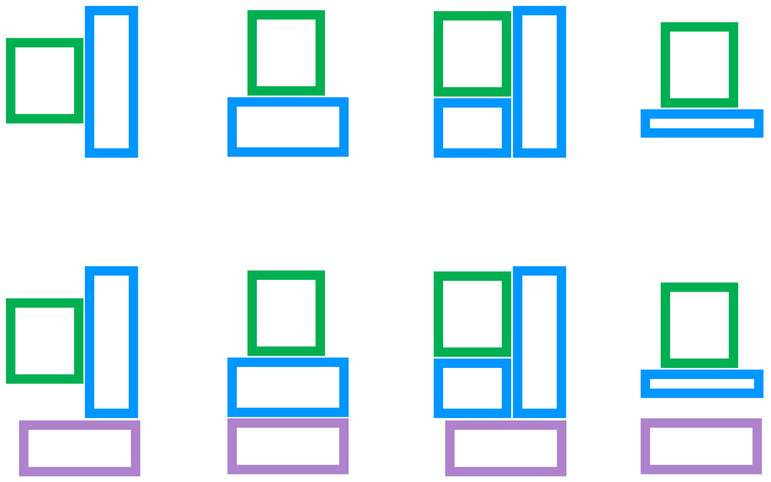
ㅡ

종성

ㄱ ㄲ ㄳ ㄴ ㄵ ㄶ ㄷ ㄹ ㄺ ㄻ ㄼ ㄽ ㄾ ㄿ ㅀ ㅁ ㅂ ㅄ ㅅ ㅆ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ

조합형 Font Template 아이디어

초성 중성 종성



19 x 4 = 76

19 x 4 = 76

총 152 자

혹시 추가적인 성능 개선의 여지는 없을까요?

비슷한 폰트로 시작하자

비슷한 폰트로 시작하자

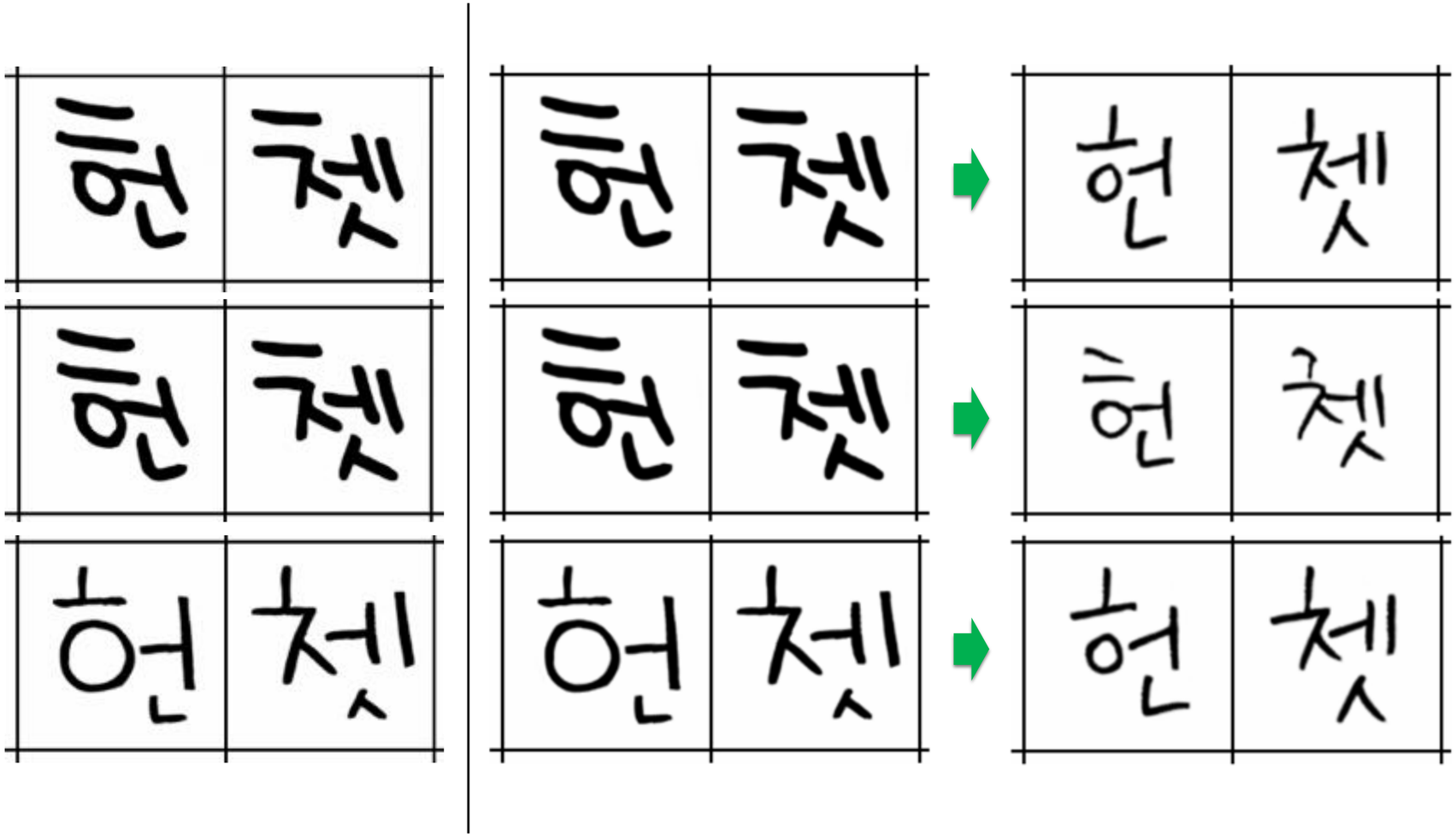
DEVIEW 2019

Metric Learning

고정 시드 글꼴 Finetune

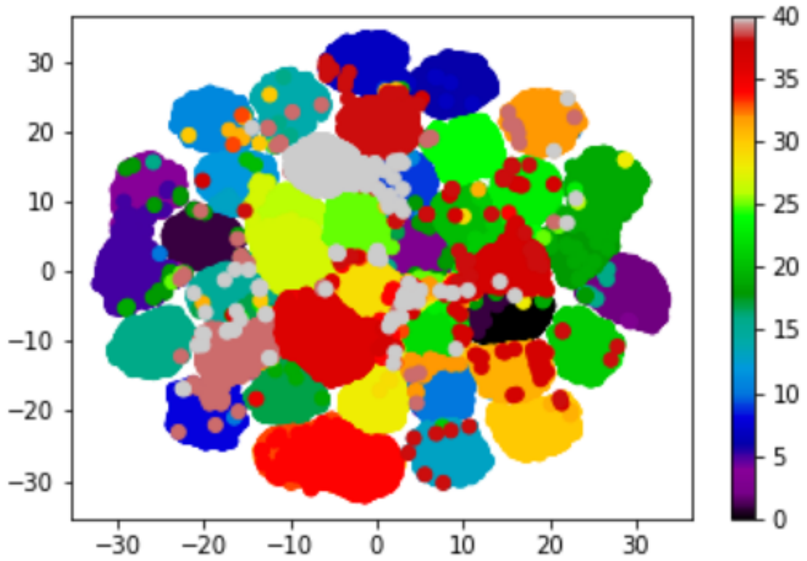
고정 시드 글꼴 Finetune

유사 시드 글꼴 Finetune



비슷한 폰트로 시작하자

Metric Learning



Query

AaBbCcDdEdFfG 안녕하세요, 폰트가 얼마나 비슷한지 테스트 하는 스트링 입니다.

Similar

AaBbCcDdEdFfG 안녕하세요, 폰트가 얼마나 비슷한지 테스트 하는 스트링 입니다.

Dissimilar

AaBbCcDdEdFfG 안녕하세요, 폰트가 얼마나 비슷한지 테스트 하는 스트링 입니다.

이제 폰트 나오나요?

아니요. 후처리해야죠.

후처리

배경의 칸 하나하나가 생성된 글자

Vector Tracing

TTF generation

Normalization

후처리

Vector Tracing

http://potrace.sourceforge.net/

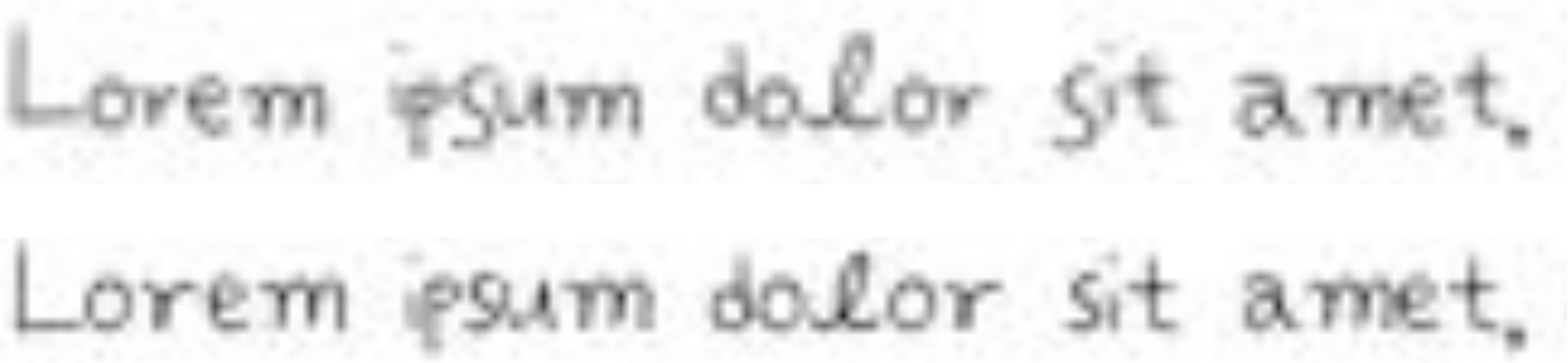
후처리

TTF generation

https://fontforge.github.io/en-US/

후처리

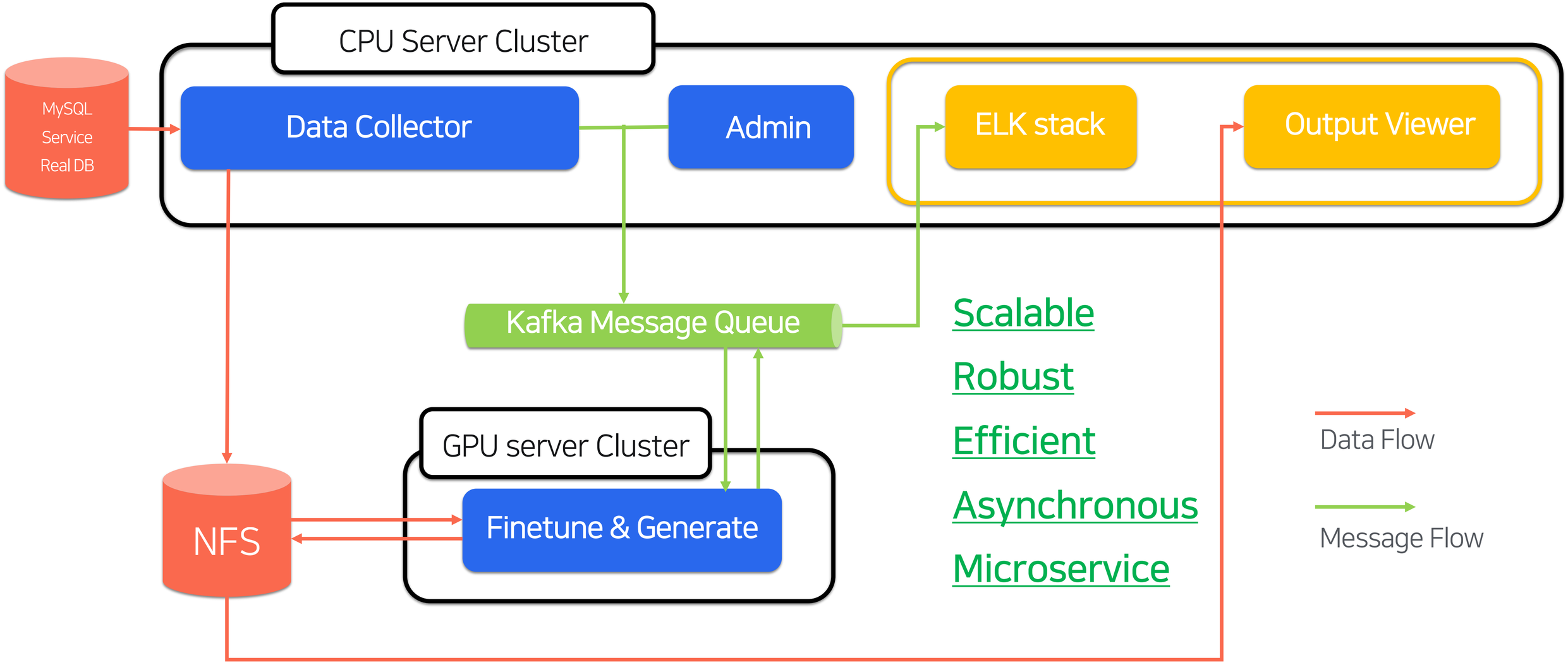
Normalization



Production Pipeline

Last but not the least.

Production Pipeline



Scalable

Robust

Efficient

Asynchronous

Microservice

한글날 공모전

이 정도로 흥할 줄은 몰랐어요

한글날 공모전

공모전 지원 이틀 차

sanghyuk.c (Clova / 전상혁) Sep 5th at 3:13 PM

참고로 말씀드리면, 지금까지 export 성공한 유저는 총 306명입니다. 의지치인 1만명까지 갈 수 있을지는 모르겠지만, 최소 2000명 정도는 모일 것 같아요

한글날 공모전

그런데...

한글날 공모전

9월 4일 부터 24일까지

20일간 25,000 명 이상 지원!!!!!

한글날 공모전 https://clova.ai/handwriting/

DEVIEW 2019

2019년 나눔손글씨 글꼴

마음을 전하는 손글씨 글꼴이 더욱 다양해지고, 손글씨에 담긴 이야기가 더욱 널리 퍼져나가길 바라봅니다.
앞으로 쓰여질 우리의 이야기를 응원합니다.

나눔손글씨 딸에게 엄마가

옛날에는 어린장에 딸 말을 써서 냉장고에 붙였던 추억이 있어요. 이제는 다 커버린 딸이 어쩌면 잊었을지 모르는 엄마의 글씨체

나눔손글씨 세아체

방고 초등학교 2학년 2반 3번입니다. 엄마 아빠가 회사에서 제 정성스러운 글씨체로 행복하게 일했으면 좋겠습니다.

나눔손글씨 무궁화

대한민국이라는 이름만 들어도 가슴이 벅찬 경험 글씨체입니다. 무궁화 꽃망울처럼 힘찬 진심을 영원히 대한민국 국민이고 싶습니다.

나눔손글씨 암스테르담

사랑하는 이가 한글을 배우고 있어요. 읽기 쉽게 자음과 모음을 써줍니다. 언어와 문화를 넘는 글씨를 나누고 싶어요.

나눔손글씨 성실체

말하지 않아도 나를 나타내는 글씨. 글씨에는 내가 담겨있고, 누군가를 향한 진심을 표현할 수 있는 가장 좋은 도구인 것 같아요.

나눔손글씨 외할머니글씨

매일 한글을 적으며, 잊지 않으려 하시는 92세 외할머니 글씨입니다. 한글을 사랑하는 외할머니의 마음을 공유합니다.

나눔손글씨 할아버지의나눔

할아버지가 써 보는 한글입니다. 요즘 유행하는 글씨체는 아니었지만 누군가에게는 꼭 필요한 글씨체였으면 좋겠습니다.

나눔손글씨 아빠글씨

직업이 운전인 아빠는 손이 많이 거치십니다. 그 손으로 첫 글씨를 쓰시는 우리 아빠! 첫 아빠 인생 응원하고, 사랑합니다.

나눔손글씨 다행체

더 정성스러운 마음을 전할 때 손글씨를 끄적입니다. 자음마다 마음의 씨앗이 담기면 좋겠어요. 당신에게 닿아 꽃이 피도록.

ㄱ-ㄷ

가람연꽃
강쫀글
강부장님체
강인한 위로
고딕 아니고 고딩
고려글꼴
곰신체
구리의 일기
금은보화
기쁨밝음
김유이체
꽃내음
꼬르마리체
나는 이거선다
나무정원
나의 아내 손글씨
노력하는 동희
노잼노잼체
다시 시작해
다진체
다채사랑
다행체
달의궤도
대광유리
대한민국 명시체
동희도박
동근인연
따뜻한 작별
따따단단
딸에게 엄마가
또박또박

ㄹ-ㅁ

마고체
맛있는체
몽돌
무궁화
무진장체
미니 손글씨
미래나무
바른정신
바른히피
반짝반짝 별
배은체체
백의의 천사
버드나무
범솔체
부장님 논지체
북극성
비상체
뽕구니깡 손글씨

ㅅ-ㅇ

사랑해 아들
상쾌한미체
성실체
세계적인 한글
세아체
세화체
소미체
소방관의 기도
손편지체
수줍은 대학생
시우 귀여워
신혼부부
아기사랑체
아름드리 꽃나무
아빠글씨
아빠의 연애편지
아인장 손글씨
아름다 자유
안성체
암스테르담
아근하는 김주임
아재장수 백금체
엄마사랑
영경쿠체
여름글씨
연지체
열아홉의 반짝임
열일체
예당체
예쁜 인경체
옥비체
와일드
외할머니글씨
윤손집이도 예뻐
우리장 손글씨
유니 빵빵빵빵
의미있는 한글

ㅈ-ㅋ

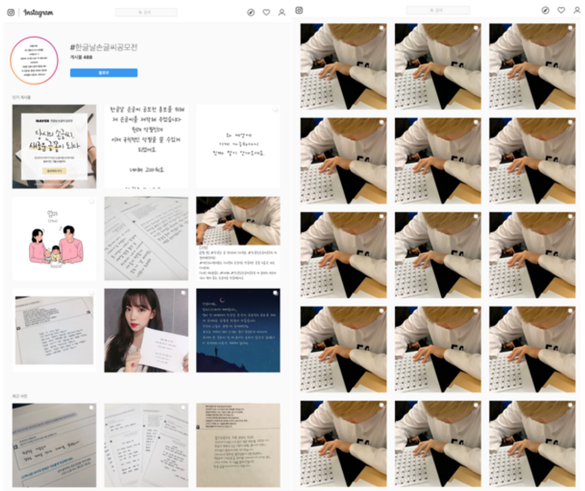
자부심지우
잘하고 있어
장미체
정꼴체
정은체
중학생
진주 박경아체
찰밥글씨
초딩희망
칠국수
코코체

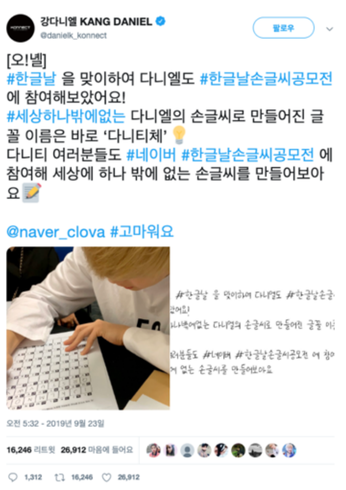
ㅌ-ㅎ

하나되어 손글씨
하나손글씨
하람체
한문체
할아버지의나눔
행복한 도서
혁이체
혜준체
호남 놀 화이팅
희망누리
흰모리수리
힘내라는 말보단

한글날 공모전

공모전 널리 알려주신 분들





벌하랑 벌랑해요>ㄷ❤

점점 추워지는 날씨 따뜻하게

잘 챙겨 입고 다니기!!

안녕하세요

우리 몰화아가씨 도현양을

사랑합니다 :)

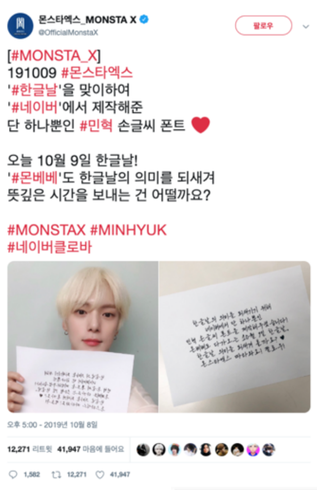
손글씨를 인사를 드리는 건 처음이네요!

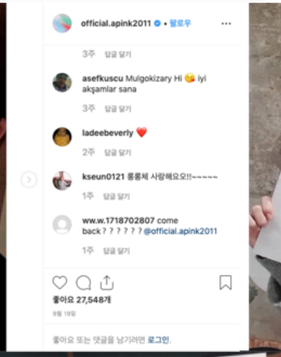
네이버에서

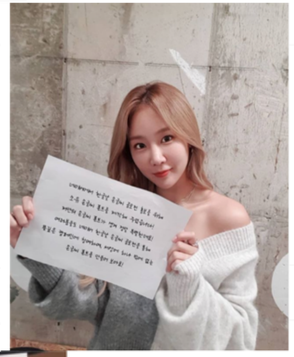
한글날 손글씨 공모전 홍보를 위해

제 손글씨를 폰트를 제작해 주셨어요

여러분들의 손글씨는 어떤가요?







네이버에서

한글날 손글씨 공모전 홍보를 위해

제 손글씨를 제작해 주었습니다

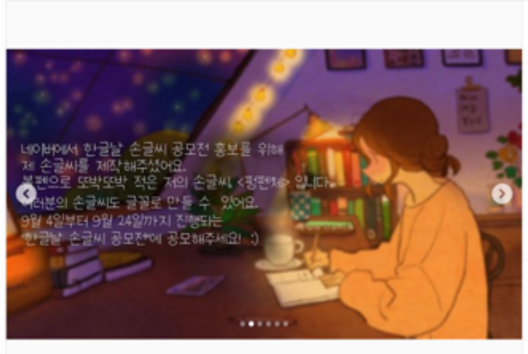
원래 악필인데

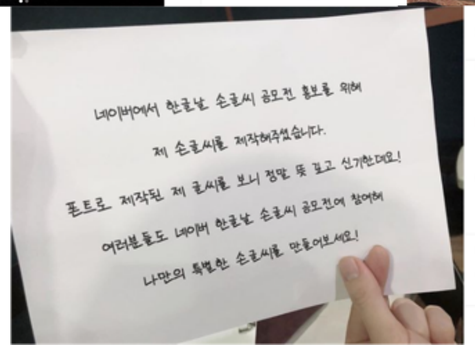
이제 규칙적인 악필을 쓸 수있게

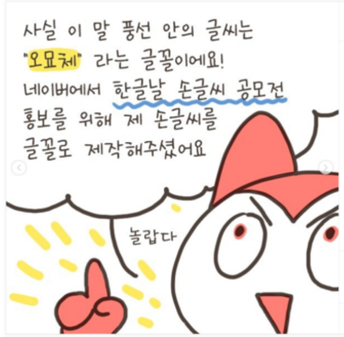
되었어요

네이버 고마워요

한글날 고마워요







공모전 에필로그 영상

What's Next?

Few Shot

현재 상황

10000 글자?

2500 글자?

400 글자?

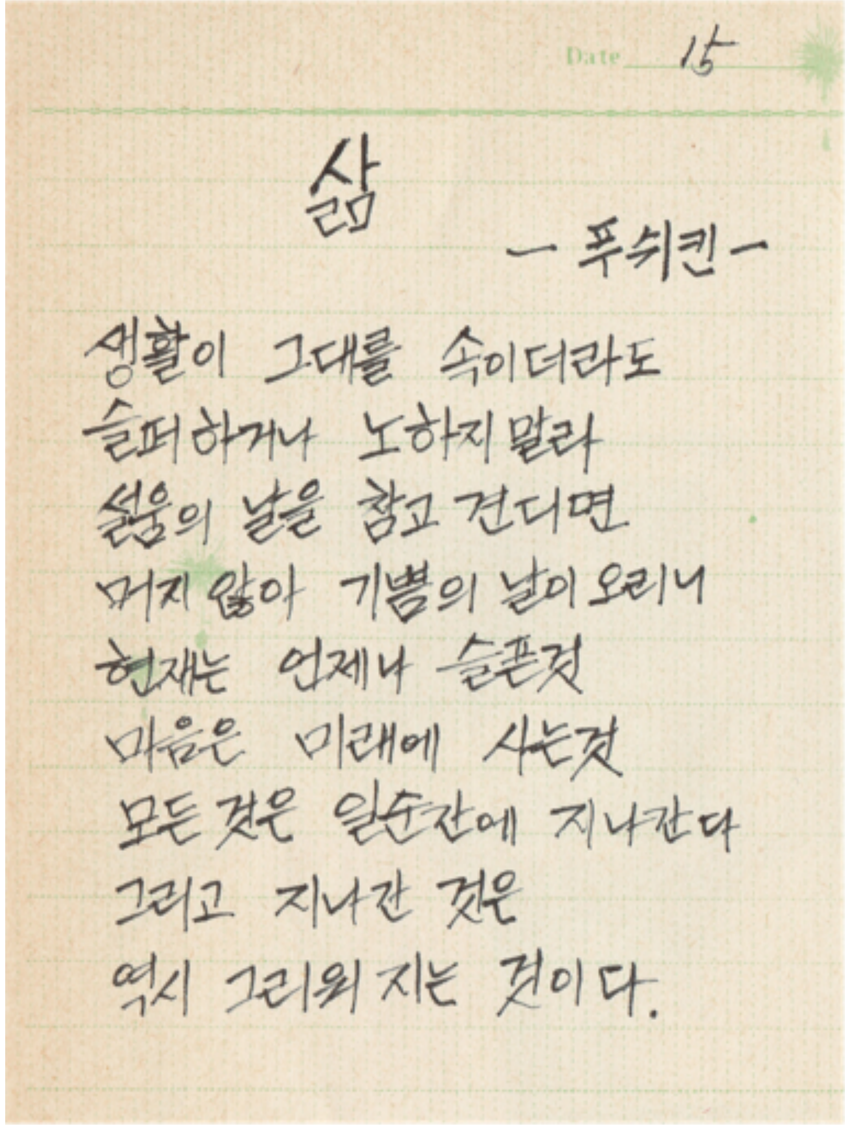
200 글자?

100 글자?

50 글자?

10 글자?

Template-less



삶
-푸쉬킨-

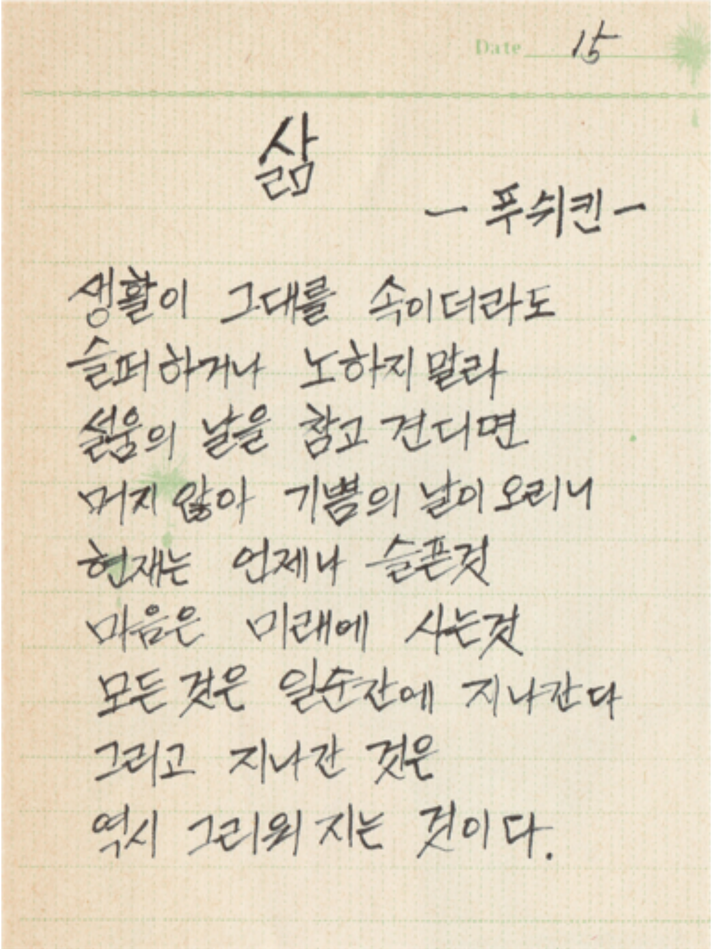
생활이 그대를 속이더라도
슬퍼하거나 노하지 말라
설움의 날을 참고 견디면
머지 않아 기쁨의 날이 오리니
현재는 언제나 슬픈것
마음은 미래에 사는것
모든 것은 일순간에 지나간다
그리고 지나간 것은
역시 그리워지는 것이다.

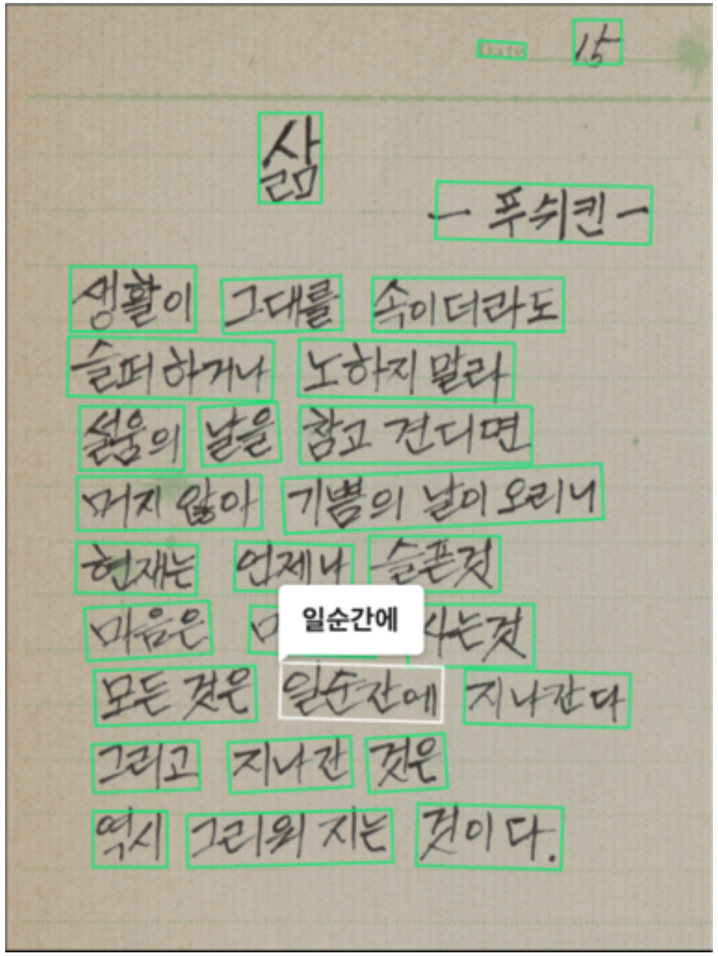
너와 당신
-푸쉬킨-

그녀가 우심코 당신이라는 공허한 호칭을
여보라는 친근함으로 불렀을 때
어리둥절해진 내 마음은
온갖 행복한 꿈들을 유발시켰다.
나는 그녀 앞에 서서 시선을 못 박고
깊은 상념에 잠겨
당신은 참 사랑스러워라고 말한다.
또한 진실로 그녀를 사랑한다고 느낀다.

손글씨 인식

https://demo.ocr.clova.ai





IEAM

高樹蔭陰 獨木不林

고수미음 독목불림

夫 우뚝만 크는 나무는 그늘을 만들지 못하고

홀로 서있는 나무는 숲을 이루지 못한다

손글씨 프로젝트 팀 소개

이바도
Prototype
Pre & Post Process
Project Manager

차준범
Modeling
Few-shot Learning
Pre & Post Process

이활석
Team Leader

전상혁
Modeling
Production Infrastructure
Code Integrator

김성현
Modeling
Model Improvement

이가영
공모전 Post Process
일본어 Pre Process

Q&A

Thank You