

# Diffusion Model을 활용한 이미지 기반의 손글씨 폰트 생성

양의종, 김다은, 조영원, 최재영  
승실대학교 컴퓨터학부

calzone0404@soongsil.ac.kr, adune@soongsil.ac.kr, ywjo@soongsil.ac.kr, choi@ssu.ac.kr

## Image Based Handwriting Font Generation Using Diffusion Model

Uijong Yang, Daeun Kim, Youngwon Jo, Jaeyoung Choi  
Dept. of Computer Science and Engineering, Soong-Sil University

### 요 약

스마트폰, PC, 태블릿과 같은 디지털 기기에서 문자를 시각적으로 표현하기 위해서는 디지털 폰트가 필수적이다. 다양한 디자인 요소가 가미된 폰트는 동일한 문장이라도 각기 다른 분위기를 연출할 수 있으며, 개인의 개성을 표현하기 위해 손글씨 폰트에 대한 관심이 증가하고 있다. 일반적으로 정형화된 폰트를 제작하기 위해서는 폰트 디자이너가 모든 글자를 직접 그릴 수 있어야 하는 과정이 필요하므로 많은 시간과 노력이 요구된다. 특히 손글씨같은 비정형화된 폰트 제작은 더욱 어렵다. 본 논문에서는 Diffusion Model 기반의 DK-Font를 활용하여 편지와 일기에 있는 손글씨 이미지로부터 한 벌의 손글씨 폰트를 자동으로 생성하는 방법을 제안한다. 또한, 손글씨 폰트가 생성된 결과를 비교하였으며, 평가지표로 그 효과성을 검증하였다.

### 1. 서론

스마트폰, PC, 태블릿과 같은 디지털 기기에서 문자를 시각적으로 표현하기 위해서는 디지털 폰트가 필수적이다. 다양한 디자인 요소가 가미된 폰트는 디지털 콘텐츠나 인쇄물에서 중요한 역할을 하며, 같은 문장이라도 전혀 다른 분위기를 연출할 수 있다. 이러한 이유로 폰트는 단순한 문자 표현을 넘어, 개성이나 감성을 전달하는 중요한 디자인 요소로 활용된다.

최근 손글씨 폰트에 대한 관심도 높아지고 있다. 손글씨 폰트는 기존의 활자 스타일의 폰트와 달리, 개인의 개성과 감성이 담긴 자유로운 형태를 갖는다. 이를 통해 디지털 환경에서도 아날로그 감성을 전달할 수 있으며, 특히 브랜드 마케팅, 캘리그래피, SNS 콘텐츠 제작 등에서 활발하게 사용된다. 그러나 폰트 제작은 폰트 디자이너가 모든 문자를 직접 그릴 수 있기 때문에 많은 시간과 노력이 요구되며 손글씨 폰트같은 비정형화된 폰트제작은 더욱 어렵다.

본 논문에서는 Diffusion Model 기반의 DK-Font를 활용하여, 편지와 같이 자유로운 형태로 작성된 글자 이미지로부터 폰트를 생성하는 방법을 제안한다. 사용자가 자유롭게 작성한 손글씨를 스캔하거나 촬영하여

생성된 이미지를 PaddleOCR[1]을 이용하여 문자를 추출하고, 이를 전처리한 후 고품질 이미지만을 선정하는 검증과정을 거친다. 검증된 문자 이미지는 DK-Font모델에 입력되어 동일한 스타일의 전체 폰트 이미지를 생성하고, 이를 폰트파일로 변환하여 사용자에게 제공한다.

### 2. 관련 연구

최근 딥러닝 기반의 이미지 생성 모델이 발전함에 따라 폰트 디자인 분야에서도 이를 활용한 연구가 활발히 진행되고 있다. 특히 다양한 스타일의 폰트를 자동으로 생성하는 기술이 주목받고 있으며, 이를 구현하기 위해 여러 생성 모델이 사용되고 있다.

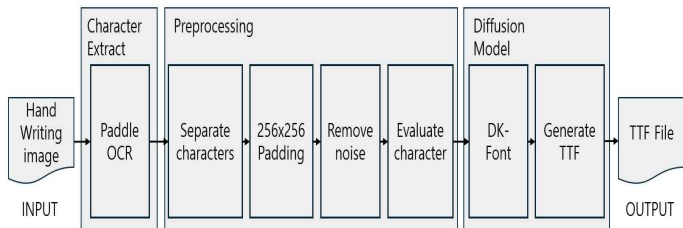
GAN(Generative Adversarial Networks)[2]은 생성자(Generator)와 판별자(Discriminator)가 서로 경쟁하며 학습하는 구조로, 이미지를 빠르게 생성할 수 있다는 장점이 있다. 하지만, 훈련 과정이 불안정할 수 있으며, 모드 붕괴(mode collapse)와 같은 문제가 발생할 가능성이 있어 안정적인 학습이 어려울 수 있다.

그러나 Diffusion Models[3]은 노이즈를 점진적으로 제거하면서 데이터를 복원하는 방식으로 동작하며, 점

진적인 변화를 통해 보다 안정적이고 높은 품질의 이미지를 생성할 수 있다는 장점을 가진다.

본 연구실에서는 선행연구로 GAN을 기반으로 한 CKFont2[4]를 개발하여, 선정된 글자를 입력하면 동일한 스타일의 전체 폰트를 생성하였다. 또한 확장된 응용 가능성을 위해 Diffusion Model을 기반으로 한 DK-Font[5]를 발표하였다. DK-Font는 소량의 스타일 이미지(Few-Shot)를 입력받아 해당 스타일을 유지하면서 전체 문자를 생성할 수 있도록 설계되었다. 모델의 세부적인 구조 및 성능 분석에 대한 내용은 향후 별도의 연구를 통해 보다 상세히 다룰 예정이다.

### 3. 손글씨 폰트 생성 방법

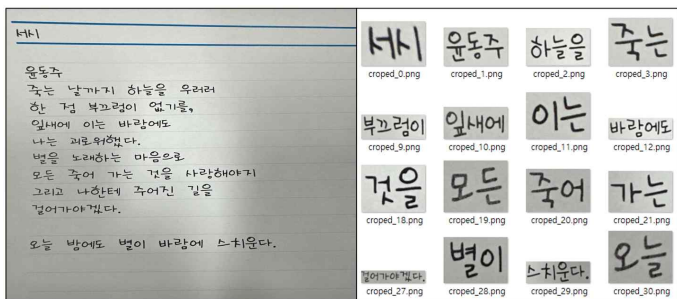


(그림1) 손글씨 생성 파이프라인

한 벌의 손글씨 폰트 생성을 위해 그림1과 같은 파이프라인을 거친다. 사용자는 자유롭게 작성된 손글씨 이미지를 입력하면 자동화 파이프라인을 거쳐 한 벌의 폰글씨 폰트가 생성된다. 일련의 과정은 Django와 FastAPI를 이용하는 End-to-End 모델로, 사용자의 개입 없이 자동으로 진행된다.

#### 3.1 문자 추출

입력된 손글씨 이미지에서 개별 문자를 추출하기 위해 PaddleOCR과 OpenCV를 이용한 문자 검출 및 전처리를 수행한다. 먼저 PaddleOCR을 사용해 그림 2와 같이 입력 이미지 내 텍스트 블록에 대한 bounding box를 검출 및 추출하여 저장한다.



(그림2) 손글씨 추출 결과(일부)

#### 3.2 이미지 전처리 및 검증



(그림3) 개별 이미지(좌) 전처리 결과 이미지(우)

추출한 문자 블록 내에는 여러 개의 문자가 포함되어 있으므로 vertical projection을 기반으로 한 박스 내에 존재하는 문자들을 개별 문자 단위로 분리한다. 이때 세로 방향으로 각 열의 픽셀 합을 계산하고, 특정 임계값 이하인 연속 구간을 기준으로 문자 간 경계를 나누어 그림3의 왼쪽과 같이 자유롭게 작성된 손글씨 이미지에서 개별 문자 이미지로 추출한다.

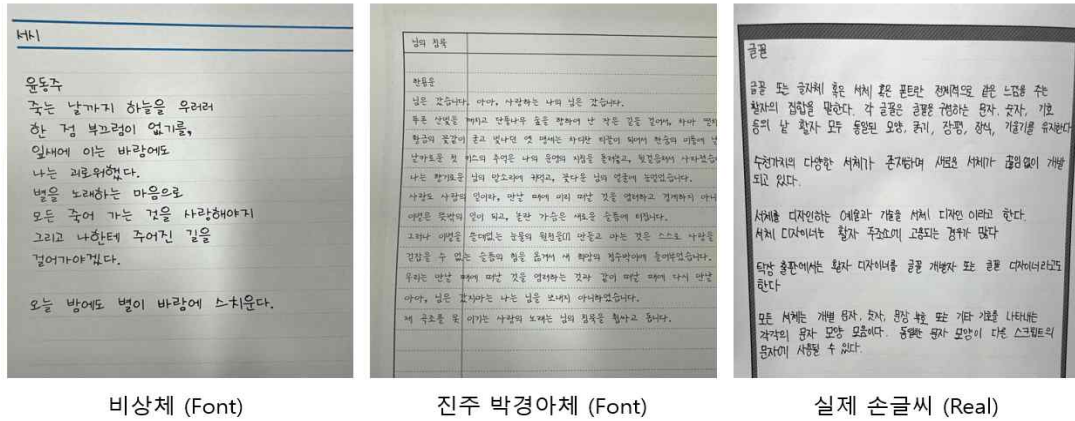
추출된 개별 문자 이미지는 사용자가 입력한 손글씨 이미지의 크기와 품질에 따라 노이즈가 포함되어 있거나 글자 크기도 일정하지 않다. 해당 이미지를 이용하여 폰트를 생성할 경우 스타일이 제대로 반영되지 않거나 문자의 의미가 사라진 이미지가 생성될 수 있다.

이를 방지하고 정상적인 폰트 생성을 위해 추출된 이미지를 256x256 크기로 Padding을 적용하고, 연결 성분의 영역 크기를 계산하여 최소 면적 기준 이상인 성분만 보존함으로써 연결 성분 기반의 노이즈를 제거하였다. 그림3의 오른쪽 이미지는 정제된 고해상도의 이미지를 보여준다.



(그림4) 오류 문자 검출 결과(일부)

일련의 과정을 마친 문자 이미지들은 Diffusion 모델 학습의 안정성과 생성 품질을 높이기 위해 문자가 원본 형태를 정확히 유지하는지를 검증한다. 검증 과정에서 왜곡되거나 불완전한 형태를 가진 오류 문자는 그림4와 같이 검출되며 모델에는 완전한 형태를 가진 고품질의 문자 이미지만이 입력된다.

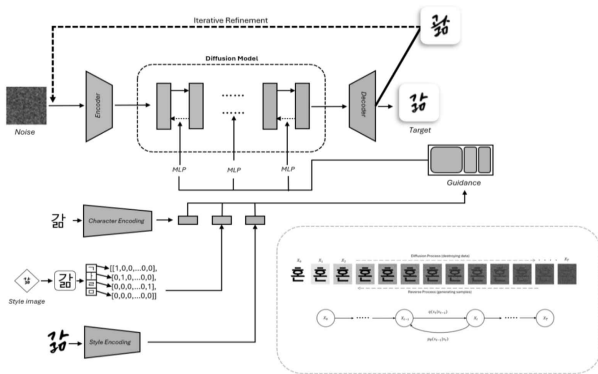


(그림6) 입력 손글씨 이미지

구분	폰트 생성 결과
나눔손글씨 비상체	이미지를 이용해 생성한 폰트입니다 다람쥐 헌 쳇바퀴에 타고파
나눔손글씨 진주 박경아체	이미지를 이용해 생성한 폰트입니다 다람쥐 헌 쳇바퀴에 타고파
실제 손글씨	이미지를 이용해 생성한 폰트입니다 다람쥐 헌 쳇바퀴에 타고파

(그림7) 생성된 폰트를 MS Word에서 사용한 결과

### 3.3 Diffusion 기반 폰트 생성



(그림5) DK-Font Architecture

검증된 문자 이미지는 DK-Font에 입력되어 전체 한 벌의 폰트 이미지를 생성한다. DK-Font는 그림 5와 같은 구조를 가지며, 글자 자체 정보(Character), 자소 구성 정보(Stroke), 스타일 정보(Style)를 각각의 인코더에서 추출한 후, 이를 통합하여 하나의 임베딩 벡터를 생성한다.

생성된 임베딩 벡터는 Denoising 과정에서 각 계층

에 주입되어 세부적인 스타일 특성을 효과적으로 전이한다. 이를 통해 소량의 스타일 이미지(Few-Shot)만으로도 동일한 스타일이 유지된 전체 폰트 세트를 생성할 수 있다. 생성된 이미지는 파이썬 라이브러리에서 제공하는 FontTools를 이용하여 한 벌의 폰트 파일을 완성한다.

### 4. 실험

일반적인 손글씨 생성 서비스에서는 전체적인 한 벌의 폰트가 없으므로, 실험에서는 기존에 있는 손글씨 폰트와 함께 사용자가 직접 작성한 손글씨를 사용하였다. 사용된 입력 이미지는 그림 6과 같이 네이버에서 제공하는 나눔손글씨 폰트 "비상체"와 "진주 박경아체", 그리고 사용자의 실제 손글씨를 포함하였다.

MS Word에서 손글씨 폰트를 사용하여 작성한 문장을 프린터로 출력하고, 스마트폰 카메라로 촬영하여 입력 이미지로 사용하였다. 파이프라인을 통해 폰트를 생성한 뒤, 생성된 폰트 이미지와 원본 폰트 이미지 간의 정성적 평가와 정량적 평가를 진행하였

다. 정성적 평가는 생성된 폰트와 기존 폰트를 비교하여 시각적인 품질과 일관성을 비교하였고, 정량적 평가는 이미지 간 비교에서 널리 사용되고 있는 평가지표인 L1 loss, SSIM, LPIPS, FID를 사용하였다.

실험 결과, 시각적인 평가로 그림7과 같이 기존 폰트의 특징을 잘 보존하면서도, 전체 문자에서 일관된 품질을 유지함을 볼 수 있었다. 이는 본 연구에서 제안하는 폰트 생성 방법이 실제 손글씨의 특징을 바탕으로 폰트를 생성하는 데 효과적임을 보여준다.

<표1> 평가지표

폰트명	L1 loss(↓)	SSIM(↑)	LPIPS(↓)	FID(↓)
비상체	19.26	0.81	0.17	48.15
진주 박경아체	16.11	0.81	0.17	53.72
평균	<b>17.68</b>	<b>0.81</b>	<b>0.17</b>	<b>50.93</b>

정량적인 결과로는 표1과 같이 L1 loss의 평균은 17.68으로 생성된 이미지와 원본 이미지 간의 픽셀 단위 차이가 비교적 작음을 의미한다. 또한, SSIM의 평균은 0.81으로, 생성된 이미지가 원본과 구조적으로 높은 유사성을 가지고 있다는 것을 알 수 있다. LPIPS의 평균은 0.17으로, 신경망 기반 지각적 유사도(perceptual similarity) 관점에서 볼 때 생성된 폰트 이미지가 원본과 상당히 비슷함을 나타낸다.

마지막으로, FID의 평균은 50.93으로 평가되었는데, 이는 생성된 폰트 이미지가 원본 폰트 이미지의 스타일과 특성을 상당 부분 유지하고 있음을 보여준다.

## 5. 결론

본 연구에서는 자유롭게 작성된 이미지에서 문자를 자동으로 추출하고 정제한 후, 한 벌의 폰트 세트를 자동으로 생성할 수 있는 방법을 제안하였다. 제안된 방법은 PaddleOCR을 활용한 문자 검출, 전처리 및 검증, Diffusion 모델을 이용한 이미지 생성 과정을 자동화하여 일관된 스타일의 폰트를 생성할 수 있도록 구성되었다.

실험에서는 손글씨 폰트와 실제 손편지를 파이프라인에 입력하여 폰트를 생성한 후, 평가지표를 통해 원본 이미지와 생성된 이미지 간의 유사성을 검증하였다. 이를 통해 한글 폰트 생성 방법이 원본과 유사한 폰트를 성공적으로 생성할 수 있음을 확인하

였다.

향후에는 생성된 폰트에서 오류가 있는 문자를 검출하고, 해당 문자만을 재생성하여 완성된 폰트 파일의 품질을 향상시킬 예정이다. 또한, 숫자나 특수기호 등 다양한 문자가 포함된 범용적인 폰트 파일을 생성하는 기능을 추가하여, 보다 다채로운 폰트 파일을 제공할 계획이다.

## Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 지원을 받아 수행되었음 (2024-0-00071)

## 참고문헌

- [1] Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., Bai, Y., Yu, Z., Yang, Y., Dang, Q., Wang, H. "PP-OCR: A Practical Ultra Lightweight OCR System", arXiv, 2020, pp. 1-10. <https://arxiv.org/abs/2009.09941>
- [2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in Proceedings of the International Conference on Neural Information Processing Systems (NIPS), 2014, pp. 2672-2680.
- [3] Ho, J., Jain, A., Abbeel, P. "Denoising Diffusion Probabilistic Models", arXiv, 2020, pp. 1-10. <https://arxiv.org/abs/2006.11239>
- [4] Jangkyoung Park, Ammar Ul Hassan Muhammad, Jaeyoung Choi, "CKFont2: An Improved Few-Shot Hangul Font Generation Model Based on Hangul Composability," Journal of KIISE, 11(12), 499-508, 2022.
- [5] Abdul Sami, Avinash Kumar, Irfanullah Memon, Youngwon Jo, Muhammad Rizwan, Jaeyoung Choi, "Text-Conditioned Diffusion Model for High-Fidelity Korean Font Generation," Proceedings of ICOIN 2025: The 39th International Conference on Information Networking, 2025, pp. 6.