



PaperCleaner, an Automatic Restoration Program of Korean Documents Based on GAN Algorithm

Min-Kyu Kim¹, Hyoung-Taek Lim¹, Gyesik Lee²

¹*Department of Computer Science and Engineering, Hankyung National University*

²*School of Computer Engineering & Applied Mathematics, Hankyung National University*

ABSTRACT

This paper describes a tool, called PaperCleaner, for the automatic restoration of Korean documents. The usefulness of PaperCleaner can be recognized when there are lots of digital papers in PDF or Jpg format whose original versions should be recovered. The need for such work would be especially critical for people working in educational institutions. Teachers for example would like to erase all the personal notes written on test papers before they can be distributed to students for exercises. But currently no applications or programs are known which automatically remove personal notes and recover the original test papers. PaperCleaner does the work without any human intervention. The main content of this paper is to introduce the main idea of our tool and to demonstrate its performance. We used our tool to restore the original test papers containing questions, answers, and some personal notes. The main algorithm is based on a Generative Adversarial Network algorithm (GAN). GAN is a class of machine learning frameworks introduced in 2014. The idea of the algorithm is that two neural networks contest with each other in the form of a zero-sum game. One agent's gain means another agent's loss. PaperCleaner is very simple to use. If a document is submitted as an input to our tool, then a restored document will be produced where almost all personal notes are removed.

© 2022 KKITS All rights reserved

KEYWORDS: Restoration of examination papers, Restoration of Korean documents, GAN algorithms, Deep learning, Machine Learning

ARTICLE INFO: Received 10 November 2022, Revised 5 December 2022, Accepted 9 December 2022.

*Corresponding author is with the School of Computer Engineering & Applied Mathematics, Hankyung National University, 327 Jungang-ro, Anseong-si, Gyeonggi-do,

17579, KOREA.

E-mail address: gslee@hknu.ac.kr

1. 서론

통상적으로 종이 문서는 서류 제출용으로 많이 사용하고 제출된 문서에 보통 펜을 이용하여 설명, 주석 등을 작성한다. 그런데 글씨를 잘못 적거나 처음부터 다시 작성해야 하는 경우 원본이 없어 난감한 상황이 발생하곤 한다.

최근 딥러닝 기법들이 의료[1], 영상처리[2-4], 사진 분류[5] 등의 분야에서 광범위하게 발전하였다. 특히 딥러닝을 이용한 그림이나 사진, 문서 등의 복원 연구[6-10]가 활발하게 진행되고 있다. 하지만 문서 복원 관련 연구의 대부분은 영어 문서에 국한되었으며[11], 한글 문서 복원에 관한 연구는 찾아보기 어렵다. 이런 이유로 한글 문서 복원이라는 주제에 관심을 두고 연구를 진행하였다. 그중에서도 특히 필기와 낙서 등으로 오염된 pdf, jpg 등의 디지털 시험지 복원에 집중했다.

시험지 속의 문제를 풀고 답과 맞추어 보는 과정에서 시험지는 답과 풀이 과정을 담은 필기 등으로 채워질 수밖에 없다. 그런데 필기를 지우지 않으면 문제의 답을 인지하게 되어 다시 풀어본다는 본래의 목적을 이루지 못하는 불상사가 일어날 수 있다. 이런 이유로 디지털 문서에 포함된 필기, 낙서, 설명 등을 제거하는 데에 사용할 수 있는 프로그램에 대한 질문이 인터넷 커뮤니티에서 많이 언급된다.

하지만 언급되는 프로그램은 모두 pdf 문서에서 낙서를 지우는, 아니면 덮어씌우는 기능을 제공하지만, 어디를 지워야 하는가는 사람이 직접 찾아서 해당 기능을 실행해야 한다. 이렇듯 문제지에 포함된 문제들을 제외한 필기 등을 제거하기 위해 적지 않은 시간과 노력이 요구된다. 특히 학교, 학원처럼 많은 양의 문제지를 대상으로 해야 하는 경우 요구되는 시간과 노력은 매우 크다.

반면에 본 논문에서 소개하는 프로그램인 PaperCleaner는 pdf, jpg 등의 디지털 시험지에 포함된 손글씨, 낙서 등을 자동으로 제거한다. 즉, 프로그램에 디지털 문서를 입력하면 자동으로 설명, 풀이 등만을 제거하고 원래의 문제는 그대로 둔, 원본에 가까운 문서를 생성한다.

PaperCleaner는 딥러닝 알고리즘인 GAN(Generative Adversarial Network)[12-14]을 이용하여 구현되었으며, 소스 코드[15]는 오픈소스로 공개되어 있다.

본 논문은 ‘2022 한국지식정보기술학회 추계학술대회’에서 발표했던 ‘Restoration of Test Paper Using GAN Algorithm’ 논문[16]의 확장이며 구성은 다음과 같다. 2장에서는 필기된 한글 문서를 복원하는 모델의 구조에 대해서 살펴본다. 3장은 모델 훈련과정을 소개한다. 4장에서는 결론을 기술한다.

2. 한글 문서 복원모델 구조

복원된 문서가 결과로 나오기 위해서는 필기된 문서를 입력받아 원본 문서와 유사하도록 생성하는 모델이 필요하다.

이러한 문제를 해결하기 위해 원본과 유사한 모조 데이터를 생성하는 생성자(Generator)와 입력된 원본 혹은 모조 데이터의 참, 거짓 유무를 구별하는 판별자(Discriminator) 두 가지 모델로 이루어진 적대적 생성신경망, GAN(Generative Adversarial Network)을 사용하였다. GAN은 이러한 두 모델이 경쟁하며 학습하는 방식으로 이루어져 있다.

<그림 1>은 생성자가 image-to-image 변환 작업을 수행하는 과정을 보여준다. 이 문제에서 사용하고 있는 자동 인코더 모델은 특정 계층(layer)까지 다운샘플링(downsapling)을 수행하는 일련의 컨볼루션 계층(convolutional layer)들로 구성된다.

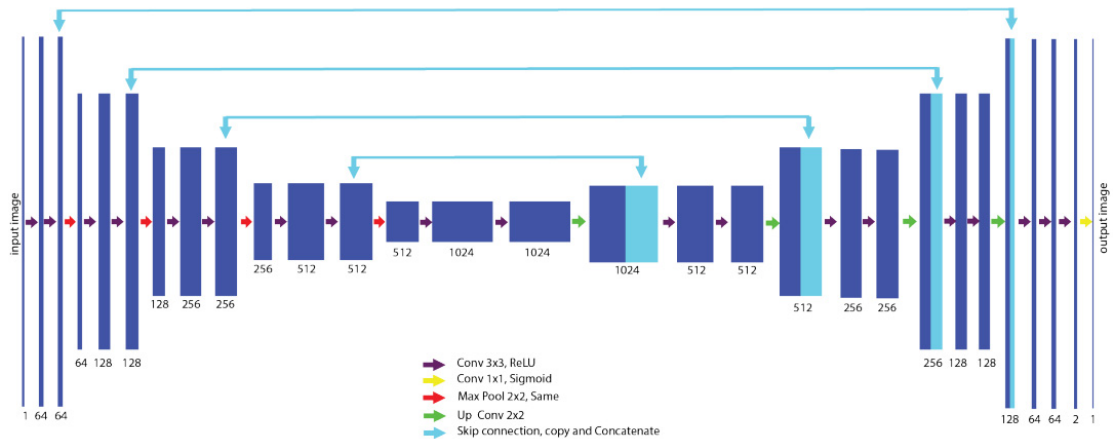


그림 1. 시험지 복원 프로그램의 생성 모델 구조

Figure 1. Structure of the Generation Model for the Test Paper Restoration Program

다운샘플링(pooling)으로 인해 많은 정보가 손실 되기 때문에 손실 전 크기의 이미지를 복구하는 것이 어렵고, 입력 및 출력이 같은 값의 픽셀을 공유하는 결과가 일어날 수 있어서 스킵 연결(skip connection)을 사용한다. 스킵 연결은 입력 데이터를 합성곱 계층(weight layer)을 건너뛰어 출력에 바로 더하는 구조로 역전과 진행 중 신호 감쇠를 막아 주기 때문에 효율적인 학습이 가능하다.

생성자는 필기가 되어있는 데이터와 문제들로만 이루어진 원본 데이터, 두 가지 데이터를 이용하여 원본 데이터와 유사한 모조 데이터를 생성해내고, 판별자는 생성된 데이터의 진위를 판단하며 생성자의 모델 성능을 향상하게 시킨다.

판별자의 모조 데이터에 대한 진위 파악 확률 분포가 0.5에 가까울수록 시험지 복원을 최상의 상태로 복원한 것이다.

판별자 모델의 경우, 진위 파악 가중치를 스스로 학습하는 데 쓰지 않고 생성자 모델로 넘겨주기 때문에 가중치 학습 기능을 제외하고 제작되어야 한다. 또한, 생성자 모델은 판별자 모델이 얻은 가

중치를 이용하여 원본과 더 가까운 모조 데이터를 만들어 낸다.

GAN 알고리즘은 아래와 같은 목적함수 $V(D, G)$ 를 이용하여 minmax problem을 계산하는 방식으로 학습하게 되며 아래 식 1과 같다.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

위 식 1에서 D 는 판별자, G 는 생성자를 의미한다. $D(x)$ 는 판별자가 원본 데이터 x 를 보고 판별하는 확률 예측값이고, $G(z)$ 는 생성자가 만들어 낸 값인 모조 데이터이다. 또한, 식의 z 를 잠재 벡터(latent vector)라고도 부르는데, 간단하게 말하자면 Noise 확률 분포로부터 추출한 Noise 샘플이다. 이를 이용하여 입력된 문서 데이터의 속성들을 학습할 수 있다. $D(G(z))$ 는 판별자가 가짜 샘플 $G(z)$ 를 보고 판별하는 확률 예측값이다. 판별자는 필기가 지워진 데이터와 원본 데이터를 비교하여 서로 같다고 판단하면 1, 아니면 0을 출력한다.

지금까지의 판별자와 생성자 모델에 대한 설명에 해당하는 훈련과정은 아래 코드로 살펴볼 수 있다.

```
real_data = clean_batch.float().cuda()
noised_data = noise_batch.float().cuda()
fake_data = generator(noised_data)
d_optimizer.zero_grad()
prediction_real = discriminator(real_data, noise
d_data)
real_data_target = torch.ones_like(prediction_re
al)
loss_real = loss1(prediction_real, real_data_targ
et)
prediction_fake = discriminator(fake_data, nois
ed_data)
fake_data_target = torch.zeros_like(prediction_r
eal)
loss_fake = loss1(prediction_fake, fake_data_ta
rget)
loss_d = (loss_real + loss_fake)/2
loss_d.backward(retain_graph=True)
d_optimizer.step()

g_optimizer.zero_grad()
prediction = discriminator(fake_data, real_data)
real_data_target = torch.ones_like(prediction)
loss_g1 = loss1(prediction, real_data_target)
loss_g2 = loss1(fake_data, real_data)*500
loss_g = loss_g1 + loss_g2
loss_g.backward()
g_optimizer.step()
```

그림 2. GAN을 이용한 한글 문서 복원모델 훈련과정 코드
Figure 2. Training Code for Restoration of Korean Document
Using GAN

<그림 2>의 코드상에서 잠재 벡터 z 를 필기된 시험지 noised_data로 사용한다. noised_data를 생성자의 입력값으로 넣어 $G(z)$ 인 fake_data를 생성한다.

판별자는 원본 데이터인 real_data와 모조 데이터 fake_data에 관한 구별을 훈련해야 하므로 각 데이터에 대한 target을 $1(\text{real_data_target})$ 과 $0(\text{fake_data_target})$ 으로 설정 후 판별자가 입력된 데이터에 대해 판별한 결과(prediction_real, prediction_fake)와 target을 입력으로 손실함수 값을 계산하여 판별자를 훈련한다.

생성자는 판별자가 fake_data에 대해 예측한 값과 그에 대한 target인 $1(\text{real_data_target})$ 을 손실함수로 계산한 값, fake_data와 real_data를 손실함수로 계산한 값을 이용해 훈련한다. fake_data가 real_data와 유사할수록 손실함수 값이 작아지므로 생성자는 fake_data를 최대한 real_data와 유사하게 만들도록 훈련하게 된다. Epoch는 80, Batch Size는 16이다. 데이터 전처리과정은 3개의 채널로 이루어진 RGB 컬러 이미지를 모델 성능의 향상과 인체, 복사 시 흑백을 사용하는 점을 활용해 단일 채널의 GrayScale 이미지로 변경하였으며, 256*256 사이즈 조각 이미지로 변형하였다.

3. 시험지 복원모델 훈련과정

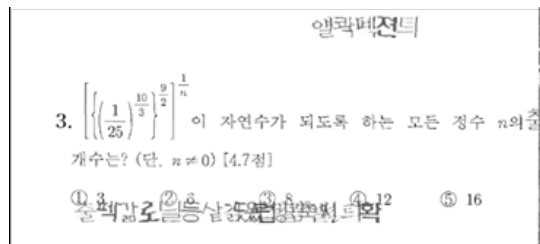


그림 3. 한글 필기가 무작위로 삽입된 이미지
Figure 3. Randomly inserted images of Korean writing

문서 복원 훈련 초반은 GAN 알고리즘을 사용하게 되면서 예상했던 것과 다르게 복원된 이미지 결과가 좋지 않았다. 많은 입력 데이터를 구하기 위해서 <그림 3>과 같이 한글 손글씨 데이터를 무작위로 삽입하여 사용했는데, 아무래도 직접 수기

로 필기 되어있는 문서와는 유사성이 없어서 그렇지 않나 추측된다.

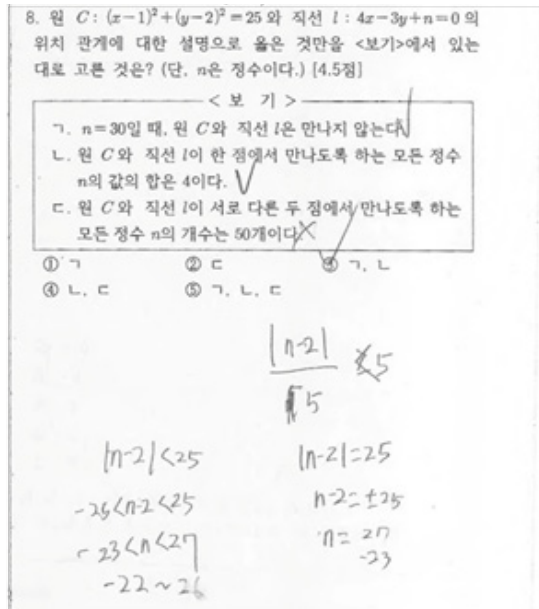
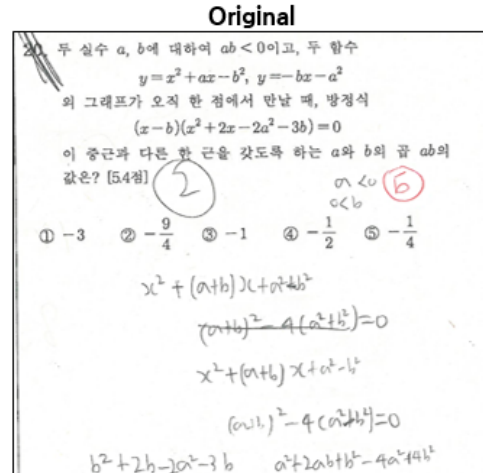


그림 4. 수기로 작성해 스캔한 이미지
Figure 4. Handwritten and scanned images

이에 따른 해결 방안으로 시험지 문서 제작 후 인쇄하여 문제를 풀었다는 가정하에 <그림 4>와 같이 직접 수기로 시험지에 필기한 뒤 다시 스캔하는 방식으로 데이터를 만들었다. 특히 무작위 필기 보다는 수식과 여러 기호가 나오는 수학 시험지를 위주로 진행하였으며, 지문이 많은 국어와 영어 시험지도 데이터에 추가되었다.

훈련 중간 결과는 초기보다 좋아졌지만 만족할 만한 수치가 나오지 않았다. 결과를 살펴봤을 때, 흐린 부분은 잘 지워지고 진한 부분은 잘 지워지지 않는 공통점을 발견했다.

그래서 연필(높은 명도)처럼 연한 부분이나 볼펜(낮은 명도)으로 생긴 진한 부분을 아래 <그림 5>와 같이 이진화(Binarization)[17]를 통해 명도를 통일시켜 입력 데이터를 개선했다.



Binarization

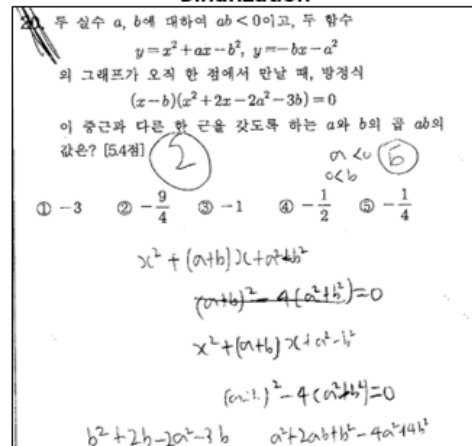


그림 5. 이진화 처리한 데이터
Figure 5. Binarized data

이진화로 명도를 통일시킨 입력 데이터로 훈련한 모델이 가장 좋은 성능을 보였다. 명도를 통일하지 않은 입력 데이터로 훈련한 모델도 준수한 성능을 보였지만, 테스트 데이터 중 일부를 제대로 복원하지 못해 일반화 성능이 떨어지는 결과가 나왔다. 아래 <그림 6>과 <그림 7>은 초기의 모델과 학습 과정을 거쳐 나온 후기 모델의 학습 성능을 보여준다.

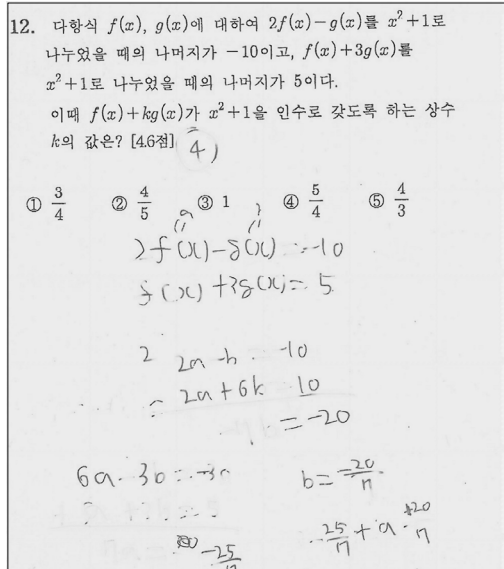


그림 6. 초기 모델 적용 시험지 데이터
Figure 6. Initial Model Application Test Paper Data

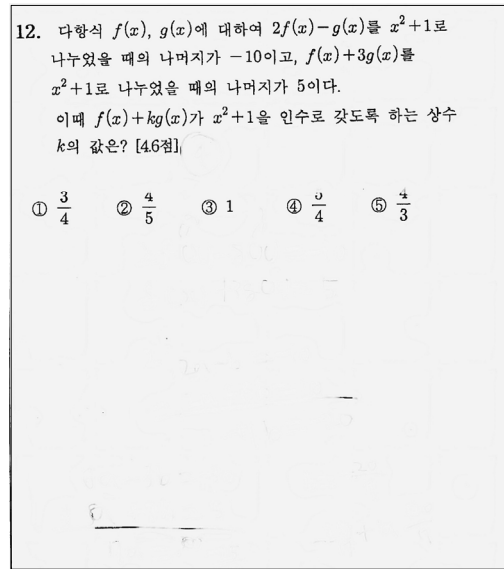


그림 7. 후기 모델 적용 시험지 데이터
Figure 7. Late Model Application Test Paper Data

로 계산한 값, fake_data와 real_data를 손실함수 계산한 값을 이용해 훈련한다. fake_data가 real_data와 유사할수록 손실함수 값이 작아지므로 생성하는 fake_data를 최대한 real_data와 유사하게 만들도록 훈련하게 된다.

2.2 학습 과정

GAN 알고리즘을 사용하면서 이론과 다르게 원된 이미지 결과는 처음부터 좋지 않았다. 추측 바로 입력 데이터로 시험지에 손글씨 데이터를 작위로 삽입하여 입력 이미지를 제작했는데, 아 래도 직접 시험 문제를 풀며 작성하는 필기와 유사성이 없어서 그렇지 않나 생각하게 되었다.

그림 8. 한글 문서 입력 데이터
Figure 8. Input Korean document data

로 계산한 값, fake_data와 real_data를 손실함수 계산한 값을 이용해 훈련한다. fake_data가 real_data와 유사할수록 손실함수 값이 작아지므로 생성하는 fake_data를 최대한 real_data와 유사하게 만들도록 훈련하게 된다.

2.2 학습 과정

GAN 알고리즘을 사용하면서 이론과 다르게 원된 이미지 결과는 처음부터 좋지 않았다. 추측 바로 입력 데이터로 시험지에 손글씨 데이터를 작위로 삽입하여 입력 이미지를 제작했는데, 아 래도 직접 시험 문제를 풀며 작성하는 필기와 유사성이 없어서 그렇지 않나 생각하게 되었다.

그림 9. 한글 문서 복원모델
Figure 9. Korean Document Restoration Model

로 계산한 값, fake_data와 real_data를 손실함수 계산한 값을 이용해 훈련한다. fake_data가 real_data와 유사할수록 손실함수 값이 작아지므로 생성하는 fake_data를 최대한 real_data와 유사하게 만들도록 훈련하게 된다.

2.2 학습 과정

GAN 알고리즘을 사용하면서 이론과 다르게 원된 이미지 결과는 처음부터 좋지 않았다. 추측 바로 입력 데이터로 시험지에 손글씨 데이터를 작위로 삽입하여 입력 이미지를 제작했는데, 아 래도 직접 시험 문제를 풀며 작성하는 필기와 유사성이 없어서 그렇지 않나 생각하게 되었다.

그림 10. DE-GAN 모델
Figure 10. DE-GAN Model

명도를 통일시킨 입력 데이터로 훈련한 모델은 기존 모델이 제대로 복원하지 못했던 테스트 데이터를 준수하게 복원하는 결과를 보여준다.

4. 한글 문서 복원모델 결과

이렇게 여러 번의 모델 훈련과정을 거쳐 복원된

시험지가 결과로 나왔으며, 이를 토대로 한글 문서 복원에 대한 데이터 전처리 과정과 모델 훈련과정을 거쳤다. 입력 데이터로 <그림 8>이 입력되었을 때, 한글 문서 복원모델과 DE-GAN[11] 모델을 거쳐 각각 <그림 9>, <그림 10>의 결과를 볼 수 있었다. 영문 문서가 바탕으로 되어있는 DE-GAN[11] 모델은 한글 문서 복원모델보다 상대적으로 복원

이 되지 않은 것을 확인할 수 있다. 아래 <그림 11>은 본 논문에서 사용한 방법을 블록 다이어그램으로 표현한 것이다.

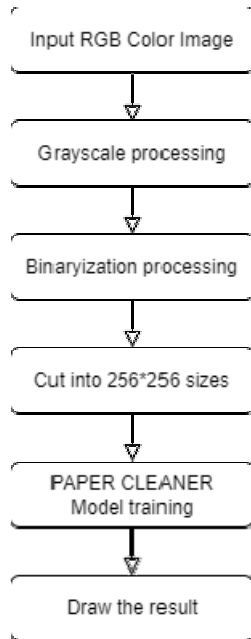


그림 11. PaperCleaner 블록 다이어그램
Figure 11. PaperCleaner Block Diagram

한글 문서 복원모델과 DE-GAN[11] 모델의 차이를 비교하기 위해 정확도와 이미지 품질 측정을 진행하였다. 각각 평균 제곱 오차(Mean Squared Error)와 최대 신호 대 잡음비(Peak Signal-to-Noise Ratio)를 구하여 <표 2>에 작성하였다.

표 2. 제작 모델과 참고 모델 mse, psnr 비교
Table 2. Comparison of mse and psnr of two models

	MSE	PSNR
한글 문서 복원모델	0.0006	80.44
DE-GAN[11] unwatermarked	0.0036	72.61

<표 2>는 동일한 한글 문서를 입력 데이터로 사용하여 제작한 한글 문서 복원모델과 DE-GAN[11] 모델로 복원한 결과이다. MSE의 경우 값이 작을수록 원본과의 오차가 적고, PSNR의 경우 값이 클수록 복원된 문서의 품질이 원본 문서와 유사할 정도로 좋다고 할 수 있다.

영문 문서 복원을 바탕으로 제작된 DE-GAN[11] 모델의 경우 PSNR 값이 72.61이 나왔으며, 이는 우리가 제작한 한글 문서 복원모델의 PSNR 값인 80.44보다 낮은 수치를 보여주었다. 이는 한글 문서에 대한 복원 능력이 DE-GAN 모델보다 본 논문에서 소개하는 한글 문서 복원모델이 뛰어나다는 것을 보여주고 있다.

5. 결 론

필기된 시험지를 입력받아 시험지 원본과 유사하도록 생성하는 모델을 통해 시험지 원본을 복원하는 프로그램을 제안했다. 이 프로그램을 통하여 시험지를 복원하면서 시간과 비용이 절감되며, 특히 학원, 학습 모임 등 시험지 원본을 소지할 수 없거나 많은 양의 시험지를 복원해야 하는 환경에서 시험지를 자주 다루게 되는 학생 및 교육 관련 종사자들에게 매우 효율적일 것으로 예상된다.

시험지 복원에 대한 데이터 전처리와 훈련과정에서 연구했던 것들을 토대로 한글 문서 복원 연구를 이어서 진행할 수 있었다. 한글 문서 복원모델의 결과는 아쉬운 성능이었지만 영문 문서 복원모델과 결과를 비교하였을 때 유의미한 결과를 보여주었으며, 특히 시험지 복원이라는 국한된 범위 내에서는 높은 성능을 보여준다.

References

- [1] S. Y. Lee, *A study of ECG classification algorithm using deep learning*, Journal of Knowledge Information Technology and Systems(JKITS), Vol. 17, No. 1, pp. 71-78, Feb. 2022.
- [2] Y. J. Lee, *A design of anomaly behavior detection system based on deep learning model using CCTV*, Journal of Knowledge Information Technology and Systems(JKITS), Vol. 17, No. 2, pp. 183-191, Apr. 2022.
- [3] K. W. Lee, S. W. Lee, H. S. Kim, and H. K. Jung, *Deep learning-based worker personal protective equipment and face identification system*, Journal of Knowledge Information Technology and Systems(JKITS), Vol. 17, No. 3, pp. 385-394, Jun. 2022.
- [4] S. M. Suh, *Deep learning based surface defect detector on metal surface*, Journal of Knowledge Information Technology and Systems(JKITS), Vol. 16, No. 5, pp. 897-904, Oct. 2021.
- [5] M. K. Park, S. J. Park, J. H. Shin, M. J. Lee, and I. S. Kim, *XAI-based dog breed classification*, Journal of Knowledge Information Technology and Systems(JKITS), Vol. 17, No. 4, pp. 785-793, Aug. 2022.
- [6] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, *Zero-shot text-to-image generation*, preprint arXiv:2102.12092, Feb. 2021.
- [7] J. U. Yun, N. Hajime, and I. K. Park, *Classification and restoration of compositely degraded images using deep learning*, JBE, Vol. 24, No. 3, May 2019.
- [8] J. W. Sin, J. H. Kim, J. Lee, C. G. Song, and S. J. Kim, *Deep learning based color restoration of corrupted black and white facial photos*, J. of KCGS, Vol. 24, No. 2, pp. 1-9, May 2018.
- [9] H. H. Jang, S. J. Ha, and G. H. Cho, *An enhancement method of document restoration capability using encryption and DnCNN*, Journal of internet of things and convergence, Vol. 8, No. 2, pp. 79-84, Mar. 2022.
- [10] J. Y. Nah, C. H. Sim, and I. K. Park, *Depth image restoration using generative adversarial network*, JBE Vol. 23, No. 5, Sep. 2018.
- [11] M. A. Souibgui, and Y. Kessentini, *DE-GAN: A conditional generative adversarial network for document enhancement*, Computer Vision and Pattern Recognition, Oct. 2020.
- [12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, Departement d'informatique et de recherche opérationnelle Université de Montréal, Jun. 2014.
- [13] Z. Wang, Q. She, and T. E. Ward, *Generative adversarial networks in computer vision: a survey and taxonomy*, ACM Comput. Surv. 54, 2, Article 37 Feb. 2021.
- [14] Q. V. Dang, and G. S. Lee, *Document image binarization by GAN with unpaired data training*, International Journal of Contents, Vol. 16, No. 2, pp. 8-18, May 2020.
- [15] PaperCleaner Source Code, https://github.com/min020/paper_cleaner, 2022.
- [16] M. K. Kim, H. T. Lim, and G. S. Lee, *Restoration of test paper using GAN algorithm*, Proceedings of the 32th KKITS Autumn Conference, Oct. 2022.
- [17] N. Otsu, *A threshold selection method from gray-level histograms*, in IEEE Transactions

on Systems, Man, and Cybernetics, Vol. 9,
No. 1, pp. 62-66, Jan. 1979.

PaperCleaner, GAN 알고리즘을 활용한 한글 문서 자동 복원 프로그램

김민규¹, 임형택¹, 이계식²

¹한경대학교 컴퓨터공학과 학부생

²한경대학교 컴퓨터응용수학부 교수

요 약

한국어 문서의 자동 복원을 위한 PaperCleaner라는 도구를 소개한다. 원본을 복구해야 하는 pdf 또는 jpg 형식의 디지털 문서가 많을 때 PaperCleaner가 유용하게 사용될 수 있다. 그러한 작업은 특히 교육 기관에서 일하는 사람들에게 중요하다. 예를 들어 교사가 학생들에게 연습용 문제지를 배포하기 전에 문제지에 작성된 모든 개인 메모를 지워야 한다. 그러나 아직은 자동으로 개인 노트 필기를 제거하고 원본 시험지를 복구하는 응용 프로그램이나 프로그램이 알려지지 않았다. 본 논문에서 소개하는 프로그램인 PaperCleaner는 이일을 자동으로 대신해 줄 수 있다. 본 논문은 PaperCleaner 프로그램의 주요 아이디어를 소개하고 프로그램의 성능을 보여준다. 예제로 PaperCleaner를 이용하여 질문, 답변 및 일부 개인 메모가 포함된 원본 시험지를 복원한다. PaperCleaner의 핵심 알고리즘은 GAN(Generative Adversarial Network) 알고리즘을 기반으로 한다. GAN은 2014년에 소개된 기계학습 프레임워크 클래스이다. GAN 알고리즘은 두 개의 신경망이 제로섬 게임의 형태로 서로 경쟁하도록 유도한다. 한 에이전트의 이득은 다른 에이전트의 손실을 의미한다. PaperCleaner의 사용법은 매우 간단한데, 복원하고자 하는 문서를 입력값으로 지정하면 거의 모든 개인 메모가 제거된 복원된 문서가 생성된다.

감사의 글

본 논문은 2017학년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2017R1D1A1B05031658)



Min-kyu Kim is a student of the Department of Computer Science and Engineering at Hankyung National University. His interests include machine learning, artificial intelligence and back-end web development.

E-mail address: kimmin020@gmail.com



Hyoung-Taek Lim is a student of the Department of Computer Science and Engineering at Hankyung National University. His interests include machine learning, artificial intelligence, and game programming.

E-mail address: aqko0719@gmail.com



Gyesik Lee received his bachelor's degree in the Department of Mathematics from Seoul National University. He received the M.S. degree and the Ph.D. degree in the Department of Mathematics and Computer Science from University of Münster in Germany. He had research positions at INRIA, AIST, and Seoul National University. He is a professor in the School of Computer Engineering & Applied Mathematics at Hankyong National University. His current research interests include machine learning, data science, and logic in computer science. He is a member of the KKITS.

E-mail address: gslee@hknu.ac.kr