탑 AI학습을 위한 최적화 기법 분석

최하진[O], 고병국[*], 이조순[*], 강은수[*], 김준오[*], 이병권[**]
[O]동국대학교 멀티미디어공학과,
[*]동국대학교 멀티미디어공학과,
[**]서원대학교 멀티미디어과
e-mail: sonic747@seowon.ac.kr

Optimization function analysis for tower AI learning

Choi Hajin[O], Ko Byeongguk[*], JoSun Lee[*], Eunsu Kang[*], Jun O Kim[*], Byongkwon Lee[**]
[O]Dept. of Multimedia Engineering, Dongguk University,
[*]Dept. of Multimedia Engineering, Dongguk University,
[**]Dept. of Multimedia Seowon University

● 요 약 ●

본 논문에서는 우리나라에 있는 탑들이 손실된 경우가 많은데 탑의 손실된 부분을 복원하기 위한 방법을 제안한다. 우리나라에 존재하는 탑들은 보존이 잘 돼있는 것보다 안 돼있는 것이 많다. 손실된 탑들을 이미지 객체로 인식시킬 시에 GAN, DCGAN, SDADE등의 알고리즘과 기존의 연구 결과들을 적용시켜 보다 효과적인 방법을 찾는 것을 제안한다.

키워드: 복원, GAN, DCGAN, SPADE GAN

I. Introduction

우리나라는 전쟁과 침략으로부터 문화재가 손실되고 회손되는 일이 많았다. 손실된 문화재 중 탑 복원에관한 연구는 미비한 상황이다. 그리고 원본에 대한 데이터가 존재하는 경우는 복원작업이 원활하게 이루어질 수 있지만, 탑의 경우 손실된 부분이 많아 원본데이터가 존재하지 않는 경우가 많다.

따라서 원본 데이터가 존재하지 않는 탑에 대해 학습하고 복원하는 알고리즘들을 비교하여 효과적인 방법을 제안하고자 한다.

II. Preliminaries

1. Related works

1.1 GAN(Generative Adversarial Network)

GAN이란 이미지를 생성하는 생성자인 Generator와 생성된 이미지를 판별하는 판별자인 Discriminator가 Adversarial하며 서로의 성능을 점차 개선해 나아가는 모델이다. 생성자인 Generator의 G와 판별자-생성자를 적대적으로 경쟁시키며 발전한다 (A)가 GAN의 기본 개념이다. Network의 N은 이 모델이 인공신경망(Aritificial Neural Network)과 딥러닝(Deep learning)으로 만들어졌다는 것을 말한다.

GAN은 그러나 단점이 존재한다. 첫 번째로 학습이 불안정하다. 또한, 생성자(Generator)가 다양한 이미지를 만들어내지 못하고 비슷한 이미지만 계속해서 생성하는 모드붕괴(Mode Collapse)현상이 나타난다.

그리고 생성자(Generator)와 판별자(Discriminator) 중 한쪽이 급격하게 발전하면 학습이 이뤄지지 않는다.

1.2 DCGAN(Deep Convolutional GAN)

위에 설명한 것처럼 학습이 어려운 GAN은 다양한 분야에서 응용되지 못했고, 수많은 실험 끝에 안정적인 학습이 가능한 GAN 모델 구조를 찾은 것이 DCGAN이다. DCGAN은 Transposed Convolution의 네트워크 구조를 가진다[4][5]. 또한, 입력데이터로 사용하는 입력 잡음(input noise)z에 대한 의미가 발견된 후, z의 값을 바꿔주면 생성 이미지가 변하는 vector arithmetic의 개념이 생겼다[1][2][3].

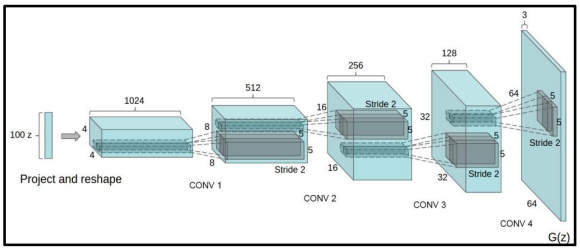

Fig. 1. DCGAN Generator Structure

그림1에 보이는 것처럼 Generator는 random input을 받아 들여 Discriminator에서 사용할 수 있는 image 데이터를 생성해내고 최종 출력물은 64x64크기의 컬러 영상이다.

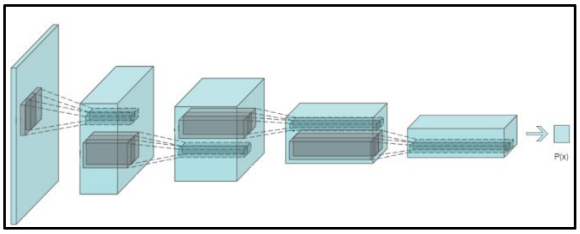

Fig. 2. DCGAN Discriminator Structure

그림2에 보이는 것처럼 Discriminator는 Generator에서 생성된 64x64크기의 컬러 영상을 입력 받아 1차원 결과를 출력한다.

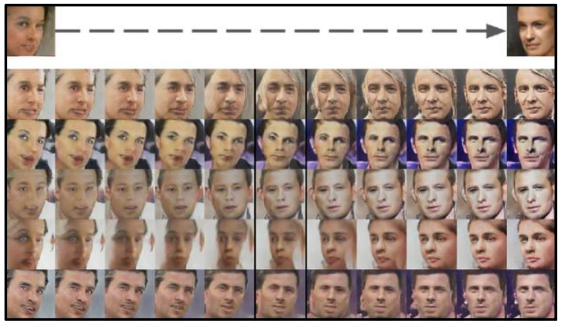

Fig. 3. DCGAN output

이렇게 학습을 통하여 얻어진 z값들을 이용하여 vector arithmetic이 가능해지고, 그림3과 같이 왼쪽과 오른쪽 얼굴에 대한 벡터들에 대한 interpolation을 수행하면 마치 얼굴의 각도가 돌아가는 영상을 얻어 낼 수 있다.

1.3 SPADE(Spatially-adaptive denormalization)

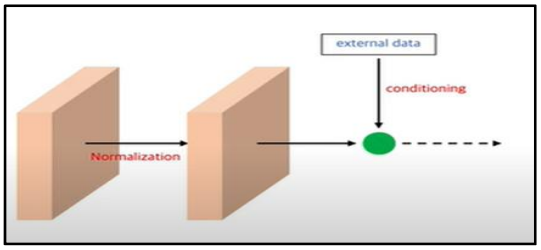

Fig. 4. Conditional Normalization Layers

기존 Normaliziotion Layer에 있어서 Image에 semantic한 부분이 convolution과정을 거치면서 손실되는 경향이 있었다. 그러나 SPADE 과정에서는 그림4에 나오는 Conditional Normalization Layer를 이용하여 convolution이 진행되어도 sematic image를 보존한다.

Conditional Normalization Layer는 기존의 Normalizion Layer와 달리 Image activation map을 normalization 하고, external data를 받아 Conditioning 시킨다. 여기서 Conditioning 방법에 따라 결과가 달라지는데 element wise로 곱셈을 해주고 다시 덧셈을 해주는 것이 SPADE 모델이다.

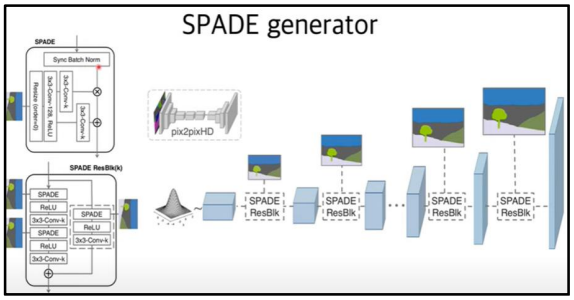

Fig. 5. SPADE generator Structure

Fig. 6. SPADE output

그림 5를 보면 random vector를 input으로 넣을 수 있기 때문에 multi-modal이 가능해진다. 또한, 보통의 encoder-decoder 구조가 아닌 decoder 구조로 resolution이 커지는 형태이고, 과정마다 spade residual block을 중간마다 넣어줌으로써, sematic map에 대한 정보를 계속 보존할 수 있다[6].

그림 6과 같이 multi-modal 합성이 가능하고 sematic image가 보존된 형태의 결과를 얻을 수 있다.

III. The Proposed Scheme

손실된 탑의 이미지 인식은 3가지 부분에 신경을 써야한다.

① 탑의 상륜부 : 탑의 윗부분

② 탑의 탑신부 : 탑의 중간부분

③ 탑의 기단부 : 탑의 기단부, 자연 암반을 기단으로 삼는 경우도

있다.

또한, 우리나라의 탑들의 경우 고려해야 하는 점들이 있다.

① 탑들 마다 문화, 시대 등의 차이로 특징들이 다르기 때문에 Labeling을 해야 한다.

② 탑들 마다 손상된 부분이 다르기에 이미지 인식 시에 특징점을 다르게 잡는 것이 효율적이라고 판단된다.

③ 또한, 배경을 삭제하여 특징점 추출을 용이하게 해야 한다.

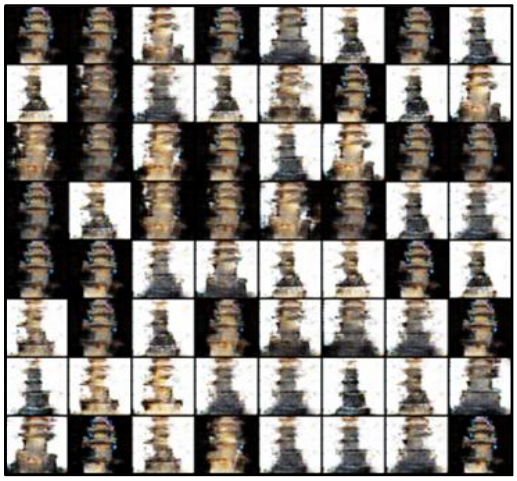

Fig. 7. Generated tower Image by DCGAN

GAN 알고리즘은 학습이 불안정할 수 있고, 풀링레이어를 사용한다. 풀링 레이어는 특징점 추출에 좋은 레이어지만 이미지의 위치 정보를 잃어버린다는 단점이 있다. 탑의 이미지 인식 시에는 위치정보가 중요하기 때문에 DCGAN을 사용하는 것이 좋다고 판단된다. 그림 7은 DCGAN을 이용하여 생성한 탑이다.

DCGAN을 이용하여 생성된 탑의 이미지를 SPADE알고리즘을 적용시킨다. 이 때, 미리 Labeling 해놓은 탑의 시대적, 문화적 특징들을 적용시키면 복원하고자 하는 탑의 시대적, 문화적 특징들이 반영되어 보다 정확한 탑 복원이 될것이라 판단된다.

IV. Conclusions

GAN 알고리즘이 발표되고 난 후, DCGAN 알고리즘의 등장으로 무수히 많은 GAN관련 알고리즘들이 나오고 있다. 이러한 알고리즘의 등장으로 이미지 복원이 가능해졌고, 2D에서 3D의 복원도 가능해졌다.

이에 본 연구는 우리나라의 문화재 중 하나인 탑의 복원에 도움이 되는 알고리즘을 연구하고자 한다.

하지만 아직 원본 이미지의 자료가 부족하고, 관련 연구들이 부족하여 어려움이 있다. 이에 본 연구는 관련 연구들과 알고리즘들 중 탑 복원에 도움이 될만한 알고리즘을 소개하여 같은 연구를 하고자 하는 사람들에게 도움을 줄 수 있다.

ACKNOWLEDGEMENT

이 (성과)는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1A2C100766811).

REFERENCES

[1] Goodfellow et al., “Generative Adversarial Nets”, NIPS, 2014

[2] Radford et al., “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, 2016

[3] Park, Taesung, et al. "Semantic image synthesis with spatially-adaptive normalization." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[4] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[5] Karras, Tero, et al. "Progressive growing of gans for improved quality, stability, and variation." arXiv preprint arXiv:1710.10196 (2017).

[6] Kupyn, Orest, et al. "Deblurgan: Blind motion deblurring using conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.