Journal of Korea Game Society 2021 Jun; 21(3): 13-24
http://dx.doi.org/10.7583/JKGS.2021.21.3.13

컴퓨터 그래픽스(공학)

3D 얼굴 모델 기반의 GAN을 이용한 게임 캐릭터 회전 기법

김한동*, 한종대**, 양희경***, 민경하**
상명대학교 일반대학원 컴퓨터학과*,
상명대학교 컴퓨터과학과**, 상명대학교 SW융합학부***
aggsae@gmail.com, {elvenwhite, yanghk, minkh}@smu.ac.kr

A GAN-based face rotation technique using 3D face model for game characters

Handong Kim*, Jongdae Han**, Heekyung Yang***, Kyungha Min**
Dept. of Computer Science, Graduate school, Sangmyung Univ.*,
Dept. of Computer Science, Sangmyung Univ.**,
Dept. of SW Convergence, Sangmyung Univ.***

요 약

본 논문은 게임 캐릭터 얼굴 일러스트레이션에 적용할 수 있는 안면 회전 기술(Face rotation) 기술을 제안한다. 기존의 진행된 연구들은 실제 사람의 얼굴 데이터에 대해서로 데이터를 한정하였으며 방대한 양의 데이터를 필요로 하였고 합성된 결과물이 좋지 못한 문제가 있었다. 본 논문에서는 기존 연구들의 존재하는 문제를 해결하기 위해 다음과 같은 방법을 도입하였다. 첫째, 입력 이미지가 갖고 있는 특징을 입힌 3D 모델을 회전시키고 다시 2D 이미지로 렌더링하여 학습 및 평가에 필요한 데이터 셋을 구축하였다. 둘째, 3D 모델을 통해 구축된 데이터에서 다양한 각도의 특징을 학습할 수 있는 적대적 생성 모델(Generative Adversarial Networks)을 설계하여 입력된 이미지를 원하는 각도로 합성할 수 있다. 논문에서는 실제 게임 캐릭터 얼굴 일러스트레이션 합성 결과를 제시한다. 합성 결과를 통해 논문에서 제안하는 방법이 잘 동작함을 확인할 수 있다.

ABSTRACT

This paper shows the face rotation applicable to game character facial illustration. Existing studies limited data to human face data, required a large amount of data, and the synthesized results were not good. In this paper, the following method was introduced to solve the existing problems of existing studies. First, a 3D model with features of the input image was rotated and then rendered as a 2D image to construct a data set. Second, by designing GAN that can learn features of various poses from the data built through the 3D model, the input image can be synthesized at a desired pose. This paper presents the results of synthesizing the game character face illustration. From the synthesized result, it can be confirmed that the proposed method works well.

Keywords : Game illustration(게임 일러스트레이션), Deep learning(심층 학습)

Received: Mar. 10. 2021 Revised: Apr. 20. 2021
Accepted: May. 01. 2021
Corresponding Author: Kyungha Min, Heekyung Yang (Sangmyung Univ.)
E-mail: minkh@smu.ac.kr, yanghk@smu.ac.kr

ISSN: 1598-4540 / eISSN: 2287-8211

Ⓒ The Korea Game Society. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.otg/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

게임은 게이머에게 더 나은 생동감과 몰입감을 제공하기 위해 3D 모델링을 한다. 하지만 3D로 캐릭터를 생성하는데 많은 비용을 필요로 한다. 한정된 자원으로 완성도 높은 게임을 만들기 위해 게임 내에서 중요하지 않은 장면(scene)들에 대해서는 3D 캐릭터를 일러스트로 만들어 게이머들에게 제공하고 있다. 그렇지만 3D 캐릭터에 비해 일러스트가 표현할 수 있는 부분이 한정적이기 때문에 게임의 생동감을 높이기 위해서 다양한 각도(pose)의 일러스트를 필요로 한다. 사람은 하나의 각도로 표현된 그림을 보고 다른 각도의 그림으로 표현하는 것에 어려움을 겪는다. 그렇기 때문에 다양한 각도의 일러스트를 만드는 작업은 많은 시간과 노력, 그리고 비용을 필요로 한다. 이처럼 많은 공수를 필요로 하는 게임 캐릭터의 다양한 일러스트를 그리는 작업에 최근 각광받고 있는 AI 기술을 도입할 경우 공수를 크게 줄일 수 있을 것으로 기대할 수 있다.

안면 회전(Face rotation) 기술은 주어진 입력 이미지의 특징을 학습하여 여러 각도로 회전시킨 결과물을 합성할 수 있는 기술이다. 이 기술은 주로 사람의 정면이 아닌 각도의 사진을 입력 받아 여러 다른 각도로 회전시킨 사람 얼굴을 합성하여 이를 통해 얼굴 인식을 하는데 주로 사용되었다. 이 기술을 게임 캐릭터 일러스트에 응용할 경우 만들고자 하는 캐릭터의 일러스트 하나와 만들기 원하는 각도를 입력할 경우 입력된 각도의 게임 캐릭터 일러스트를 얻어낼 수 있을 것으로 기대할 수 있다.

현재 안면 회전 기술은 대다수의 연구들에서 적대적 생성 네트워크(이하 GAN, Generative Adver-sarial Network)[1]을 활용하고 있다. 학습 데이터가 갖고 있는 특징들을 학습하고 학습된 특징을 바탕으로 입력 이미지에 대한 회전(rotation)을 수행하여 원하는 결과를 합성한다. 대다수의 안면 회전 기술의 적용되는 데이터는 주로 사람의 얼굴 데이터로 구성되어 있어 게임 캐릭터 데이터가 입력으로 주어졌을 경우에는 원하는 결과를 합성할 수 없다는 한계점을 갖고 있다. 직관적으로 생각했을 때, 게임 캐릭터로만 이루어진 데이터 셋을 사용하여 안면 회전 네트워크를 학습시키면 원하는 대로 동작할 수 있을 것이라는 생각을 할 수 있다. 하지만 게임 캐릭터로만 이루어진 데이터 셋이 존재하지 않는 다는 문제점에 마주하게 된다. 대다수의 안면 회전 기술들은 사람의 회전된 얼굴을 합성하는 것을 목표로 하고 있으며, 이에 따라 사람 얼굴 데이터 셋에 대한 연구가 대다수를 차지하고 있다. 사람의 얼굴 데이터 셋을 사용하였을 경우 게임 캐릭터와 사람 얼굴 간의 차이로 인하여 적절한 결과를 합성할 수 있다는 한계에 직면하게 된다. 따라서 본 논문에서는 다음과 같은 과정을 통해 언급된 한계를 극복하려고 하였다. 첫째, 게임 캐릭터로 이루어진 데이터 셋이 없다는 한계를 극복하기 위해 3D 모델을 활용하였다. 주어진 게임 캐릭터 데이터를 입력으로 주어, 해당 데이터가 갖고 있는 질감 특징(texture)을 3D 모델에 입힌다. 이를 통해 게임 캐릭터의 얼굴 특징을 갖고 있는 3D 모델이 생성된다. 생성된 3D 모델의 회전을 통해 다양한 각도의 게임 캐릭터의 모습을 얻을 수 있다. 3D 모델을 통해 회전시킨 게임 캐릭터를 다시 2D 이미지로 만든다. 하나의 게임 캐릭터 데이터는 이 과정을 거쳐 다양한 각도의 게임 캐릭터 데이터를 렌더링하여, 소수의 게임 캐릭터 데이터 셋으로 많은 게임 캐릭터 데이터 셋을 구성할 수 있다. 둘째, 안면 회전을 위해서 GAN 모델을 작성하였다. 작성된 GAN 모델은 기존에 잘 구성되어 있는 사람 얼굴 데이터 셋으로 사전 학습(Pre-trained)하여 모델이 다양한 특징을 학습할 수 있도록 구성하였다. 사전 학습된 모델에 에 첫 번째 단계에서 얻어낸 게임 캐릭터 데이터 셋을 주어 모델이 게임 캐릭터의 특징을 학습할 수 있도록 하였다. 제안된 방법을 통해 입력으로 주어진 게임 캐릭터 일러스트에 대해 회전을 수행하여 원하는 결과를 합성할 수 있다. 합성된 결과는 입력으로 주어진 데이터의 그림체나 특징을 잘 보존하고 있어 사람이 다양한 각도의 일러스트를 만들 때 들여야 하는 공수를 크게 줄일 수 있을 것을 기대할 수 있다.

2. 관련 연구

2절에서는 이전에 진행되었던 다양한 안면 회전 기술 연구에 대해서 소개한다. 안면 회전 기술을 위해 다양한 방법으로 연구되었으며 이는 심층 학습(Deep learning) 이전과 이후로 나눌 수 있다.

2.1 심층 학습 도입 이전의 안면 회전 기술

심층 학습 등장 이전에 진행된 안면 회전 기술 연구들은 Hassner[2], Zhu[3]의 연구에서 사용된 것과 같이 3D 모델을 사용하였다. 입력 이미지의 특징을 3D 모델에 입힌다. 이후 이미지의 질감이 입혀진 3D 모델을 회전하는 방식으로 안면 회전 기술을 구현하고자 하였다. 또 다른 방법으로 진행된 안면 회전 기술 연구는 Moniz[4]의 연구가 있다. Moniz의 연구에서는 3D 변환 표(3D transform matrix)를 통해 2D 와 3D의 각 포인트를 매핑하는 방식의 안면 회전 기술을 제안하였다. 이러한 연구들은 일정 수준 이상의 성과를 얻을 수 있었지만 2D 이미지를 3D로 변환하는 과정으로 인한 왜곡(Distorted) 현상과 흐릿함(Blurry) 현상이 합성된 결과물에서 발견되는 한계를 보였다.

2.2 심층 학습 도입 이후의 안면 회전 기술

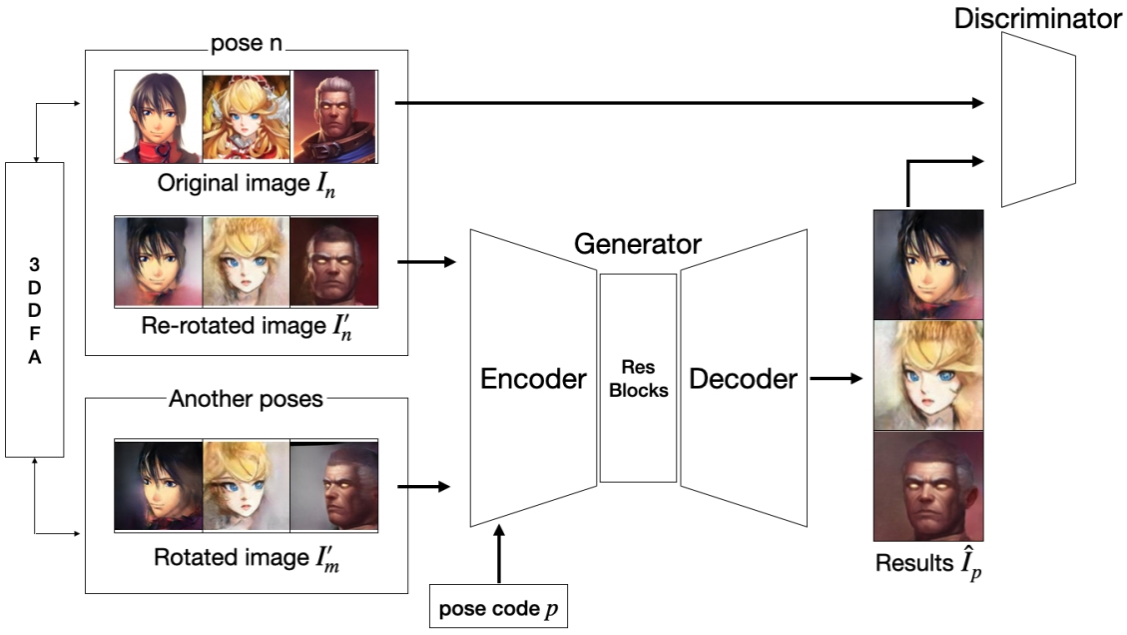
심층 학습 분야에 대한 많은 연구가 이루어짐에 따라 안면 회전 기술에도 심층 학습을 도입하려는 연구가 진행되기 시작하였다. 그 중에서도 특히 GAN을 도입한 안면 회전 기술 연구가 활발하게 진행되었다. GAN을 활용한 안면 회전 기술 연구는 크게 입력 이미지와 같은 각도의 결과물을 합성하는 재건(Reconstruction)하는 방식과 데이터가 갖고 있는3D 기하학 정보를 활용하는 방식으로 나누어져 진행되었다.

2.2.1 재건 방식의 안면 회전 기술

재건 방식은 입력 이미지와 동일한 각도의 합성 이미지를 만드는 GAN 기술을 의미한다. 주로 측면 입력 데이터를 정면 이미지로 합성해주는 정면화(Frontalization)를 목적으로 합성된 결과물은 입력 이미지가 어떤 사람의 얼굴인지 식별하는 얼굴 인식(Face recognition)을 위해 사용되었다. 대표적인 재건 방식을 활용하여 진행된 연구들에는 Tran[5], Huang[6], Hu[7], Qian[8]이 있다. Tran은 입력 이미지의 특징과 입력 이미지의 각도를 분리하여 입력 이미지의 각도와 상관없는 정면 이미지를 만드는 DR-GAN에 대해서 제안하였다. Huang은 입력 이미지의 전체적인 윤곽 특징과 눈, 코, 입과 같은 세부 특징을 따로 학습시켜 결과 이미지를 합성하는 TP-GAN을 제안하였다. Hu의 연구에서는 입력 이미지의 히트맵을 활용한 CAPG-GAN을 제안하였다. Qian의 연구에서는 라벨이 표현된 데이터와 라벨이 표현되지 않은 데이터를 합쳐 학습하는 FNM을 제안하였다. 재건 방식으로 진행된 연구들은 정면에 가까운 각도의 입력 이미지에 대해서는 대체로 좋은 정면화 성능을 보였지만 측면에 가까운 각도를 갖고 있는 입력 이미지에 대해서는 제대로 된 성능을 보이지 못하며 정면 이외의 다른 각도에 대한 결과물을 합성할 수 없다는 한계를 보였다.

2.2.2 3D 기하학 정보를 활용하는 방식의 안면 회전 기술

3D 기하학 정보를 활용하는 방식의 연구들은 심층 학습이 도입되기 이전의 연구에서 사용했던 3D 정보를 활용하는 방식과 GAN을 합친 방법을 사용하여 진행하였다. 대표적인 3D 기하학 정보를 활용한 연구들에는 Yin[9], Deng[10], Cao[11], Zhou[12]가 있다. Yin은 기존 3D 모델을 활용하는 3DDM의 계수를 네트워크로 가져와 결과 합성에 활용한 FF-GAN에 대해서 제안하였다. Deng은 UV-map을 활용한 UV-GAN을 제안하였다. Cao는 입력 이미지의 특징 정보를 합성하는 결과에 적용한



[Fig. 1] Overview

HF-PIM에 대해서 제안하였다. Zhou는 입력 이미지의 질감 특징 정보를 3D 모델에 입히고 질감 특징이 적용된 3D 모델을 2D 이미지로 다시 렌더링하여 데이터 셋을 구성하고 구성한 데이터 셋을 통해 여러 각도의 결과를 합성할 수 있는 Rotate-and-render를 제안하였다. 3D 기하학 정보를 활용하는 방식으로 진행된 연구들은 앞서 살펴본 재건 방식을 적용한 연구들에 비해서 더욱 측면에 가까운 각도의 입력 이미지에 대해서도 좋은 성능의 결과를 얻을 수 있었지만 추가로 3D 모델에 대한 정보와 처리를 필요로 하여 더 많은 자원을 소모한다는 한계를 보였다.

2.3 한계점을 극복하기 위해 사용한 방법

재건 방식과 3D 기하학 정보를 활용한 안면 회전 기술 연구들은 각각의 장점과 한계점을 갖고 있다. 두 방법의 공통적인 한계점으로는 합성한 결과물이 실제 사용자가 원하는 결과물인지 확인하기 위한 지면 진실(이하 GT, Ground truth)이 필요하다는 것이다. 이로 인해 기존에 진행된 안면 회전 기술 연구들에서는 특정 사람의 정면 이미지와 함께 해당 사람의 다른 각도의 이미지를 쌍으로 갖고 있는 데이터 셋을 활용하였다. 때문에 모델에 사용되는 이미지와 해당 이미지의 지면 진실에 해당하는 이미지를 쌍으로 갖고 있는 데이터 셋을 필요로 한다. 그렇기 때문에 이를 게임 캐릭터 일러스트를 대상으로 동작시키기 위해서는 방대한 양의 쌍으로 구성된 게임 캐릭터 데이터 셋을 필요로 하지만 해당하는 게임 데이터 셋에 대한 연구가 활발하게 이루어지지 않은 실정이다. 따라서 본 논문에서는 쌍으로 된 데이터 셋을 사용하지 않고 단일 이미지로 구성된 데이터 셋을 활용한 연구에서 영감을 받아 별도의 추가 게임 캐릭터 데이터 없이 게임 캐릭터의 얼굴을 회전시킬 수 있는 방법과 이전에 진행된 안면 회전 기술들을 위한 네트워크로 활용된 GAN에서 영감을 받아 결과 이미지를 잘 생성하기 위한 방법에 대한 연구를 진행하였다.

3. 방법론

[Fig. 1]은 본 논문에서 원하는 각도의 게임 캐릭터를 합성하기 위해 제안하는 모델의 전체 구조를 도식화한 것이다. I_n은 학습을 위해 사용되는 n의 각도를 취하고 있는 입력 이미지를 의미한다. 이 때, n은 이미지에 표현되어 있는 얼굴의 각도를 의미하며 -90° ~ +90° 사이의 값을 가진다. I'_m은 입력 이미지 I_n을 3D 모델에 적용 후 m 각도로 회전시킨 후 다시 2D로 렌더링 한 이미지를 의미한다. $\hat{I}_p$은 입력 이미지와 회전시킨 이미지를 활용하여 학습된 모델 f에서 주어진 입력과 각도에 대해 합성한 결과 이미지를 의미한다. 하나의 이미지와 원하는 각도 p를 입력하여 결과 이미지를 합성할 수 있다.

본 논문에서 제안하는 모델은 크게 두 단계로 나누어져 동작한다. 첫 번째 단계는 학습에 사용하기 위한 데이터를 구축하는 단계이다. 앞서 언급한 바와 같이 기존의 안면 회전 기술들은 합성한 결과물의 품질을 향상시키기 위해 GT 데이터를 필요로 하는, 쌍으로 이루어진 데이터 셋이 필요하다는 점을 언급하였다. 이 단계를 통해 해당 문제를 해결할 수 있다. 이를 위해 단일 이미지를 GT로 한, 쌍을 이룰 데이터들을 생성할 필요가 있다. 첫 번째 단계에서는 2D로 되어 있는 주어진 입력을 3D 모델로 만들어 주는 오픈 소스 3DDFA[14]를 활용한다. n 각도를 취하고 있는 원본 이미지 I_n을 3DDFA에 입력하면 입력 이미지가 취하고 있는 각도 n의 3D 모델을 생성하여 해당 모델에 입력 이미지가 갖고 있는 재질 특징을 입힌다. 이렇게 입력 이미지의 재질 특징을 갖고 있는 3D 모델은 임의의 각도 m으로 회전하게 되고 이를 렌더링하여 m 각도의 2D 이미지 $I_m{}'$를 얻을 수 있다. 이렇게 생성된 m 각도의 이미지에 대해 같은 방법을 적용하여 다시 n 각도를 갖고 있는 이미지 I'_n를 만든다. 이는 Cycle GAN[13]에서 이미지와 이미지 사이의 변환을 할 때 사용하는 것과 유사하게 동작한다. 이 단계에서 생성된 이미지를 활용하면 입력 이미지에 대한 여러 각도의 특징을 모델 f가 학습할 수 있다. 원본 이미지는 GT로써 모델 f가 정확한 결과물을 합성했는지 확인하는데 사용된다. [Fig. 1]의 좌측에 위치한 그림이 이런 과정을 표현하고 있다. 첫 번째 단계를 거쳐 생성된 데이터들은 이후 두 번째 단계의 모델 학습 및 평가 단계에서 학습 및 평가 데이터로써 활용된다.

두 번째 단계는 결과물을 합성하기 위해 제안된 모델을 학습하고 평가하는 단계이다. 이 때, 첫 번째 단계에서 생성한 데이터를 활용한다. 생성된 데이터는 입력 이미지의 GT와 3DDFA를 통해 생성된 입력 이미지와 같은 각도의 이미지와 랜덤하게 회전된 측면 각도의 이미지로 구성이 되어 있다. 모델은 입력 이미지와 같은 각도의 이미지와 랜덤한 측면 각도의 이미지로부터 여러 각도에 대한 특징을 학습하게 되고, 이를 통해서 입력된 데이터와 각도에 대해서 학습한 특징을 활용하여 결과물을 합성할 수 있게 된다. 모델은 인코더(encoder)-디코더(decoder) 구조를 갖고 있는 생성자(generator) G와 인코더와 유사한 구조로 되어 있는 분류자(discriminator) D로 이루어진 GAN 으로 이루어져 있다. 모델의 구조에 대한 자세한 정보는 2.3.1 절에서 확인할 수 있다. 생성자 G는 (1)의 수식과 같이 첫 번째 단계에서 두 번의 3DDFA를 적용하여 나온 이미지 $I_n{}'$와 $I_m{}'$와 합성을 원하는 각도 p를 입력 받아 p 각도를 취하는 결과물을 합성하게 된다.

$$G(I'_n, I'_m, p) = \hat{I}_p \tag{1}$$

생성자 G에서 합성한 이미지 $\hat{I}_p$는 분류자 D에서 GT I_n와 함께 입력되어 모델을 학습한다. 학습이 완료된 네트워크에 합성을 원하는 이미지와 포즈를 입력하면 해당 포즈를 갖고 있는 이미지를 합성

[Table 1] The architecture of model

Description	Generator		Discriminator
	encoder	decoder	
Initial Block	Conv(3, 64) InstanceNorm(64) ReLU(false)	None	None
Each Block Information	Conv(64, 128) InstanceNorm(128) ReLU(false)	ConvTranspose(2048, 1024) InstanceNorm(1024) ReLU(false)	Conv(3, 64) LeakyReLU(0.2, false)
	Conv(128, 256) InstanceNorm(256) ReLU(false)	ConvTranspose(2048, 1024) InstanceNorm(1024) ReLU(false)	Conv(128, 256) LeakyReLU(0.2, false)
	Conv(256, 512) InstanceNorm(512) ReLU(false)	ConvTranspose(2048, 1024) InstanceNorm(1024) ReLU(false)	Conv(256, 512) LeakyReLU(0.2, false)
	Conv(512, 1024) InstanceNorm(1024) ReLU(false)	ConvTranspose(2048, 1024) InstanceNorm(1024) ReLU(false)	Conv(512, 512) LeakyReLU(0.2, false)
ResBlock	ReflectionPad(1) Conv(1024, 1024) InsanceNorm(1024) ReLU(false) X18	None	None
Final Block	None	ReflectionPad(3) Conv(128, 3) Tahn()	Conv(512, 1)

할 수 있다.

제안하는 모델의 구체적인 정보에 대해서는 3.1절에서 확인할 수 있으며 모델을 활용하여 합성한 이미지 결과는 4.3절에서 확인할 수 있다.

3.1 모델 구조

본 논문에서 제안한 모델은 게임 캐릭터의 이미지와 합성하기를 원하는 각도를 입력받아 해당 각도를 취하고 있는 결과 이미지를 합성한다. 3.1절에서는 이를 가능하게 하는 제안된 모델에 대한 내용을 다룬다.

3.1.1 모델 정보

3.1.1절에서는 제안하는 모델이 어떻게 구성되어 있는지에 대한 정보를 제공한다. 제안하는 모델의 생성자 G는 입력된 데이터로부터 특징을 학습하는 인코더와 얻어낸 특징을 통해 합성 이미지를 생성하는 디코더 구조로 이루어져 있으며, 인코더와 디코더 사이에는 누락되는 특징을 줄이기 위한 ResNet block이 존재한다. 합성된 이미지가 GT와 유사하게 합성하였는지 파악하기 위한 분류자 D의 경우 인코더의 구조와 유사한 구조를 가지고 있으나, 각 레이어에 입력되는 매개변수(Parameter)와 사용된 활성 함수에서 차이를 보인다. [Table 1]을 통해 모델 구조에 대한 정보를 확인할 수 있다.

3.1.2 손실 함수(Loss functions)

3.1.2절에서는 제안된 모델이 GT에 가까운 결과물을 합성하기 위해 채택한 손실 함수에 대해 소개한다. 제안된 모델에서는 총 4개의 손실 함수에 실험을 통해 얻은 최적의 가중치를 더하여 활용한다.

첫 번째 손실 함수는 적대적 손실(Adversarial loss)이다. 이는 GAN에서 사용되는 손실 함수로 분별자 D를 통해 생성자 G가 만들어낸 결과 이미지 $\hat{I}_p$와 GT I_p에 얼마나 가까운 결과인지를 알아낼 수 있다.

$$L_{GAN}(G,D) = \mathbb{E}[logD(I_p)] + \mathbb{E}(log(1 - D(\hat{I}_p)] \quad (2)$$

이를 수식으로 (2)와 같이 표현할 수 있다. 적대적 손실을 통해 생성자는 분별자에서 합성한 결과물이 GT라고 판별할 수 있는 보다 더 GT에 가까운 결과물을 합성할 수 있게 된다.

두 번째 손실 함수는 특징 일치 손실(Feature matching loss)이다. 이는 입력 이미지와 합성 이미지가 분류자의 각 영역을 거치며 나오는 특징을 비교하는 손실 함수이다.

$$L_F(G,D) = \frac{1}{N_D}\sum_{i=1}^{N_D} ||D^i(I_a) - D^i(\hat{I}_a)||_1 \quad (3)$$

이를 수식으로 (3)과 같이 표현할 수 있다. 이를 통해 입력 이미지와 합성 이미지가 분류자에서의 얻어지는 특징 차이를 최소화하여 GT에 특징을 잘 보존하고 있는 결과 이미지를 합성할 수 있다.

세 번째 손실 함수는 지각 손실(Perceptual loss)이다. 이 손실 함수는 ImageNet으로 사전 학습된 VGG 네트워크를 또 다른 분류자로 활용하였을 뿐, 계산 방법은 특징 일치 손실과 동일하다.

$$L_{vgg}(G,D) = \frac{1}{N_{vgg}}\sum_{i=1}^{N_{vgg}} ||VGG^i(I_a) - VGG^i((\hat{I})_a)||_1 \quad (4)$$

지각 손실을 통해 입력 이미지의 특징과 합성 이미지의 특징을 정규화할 수 있다.

네 번째는 종합 변형 정규화(Total variation regularization)이다. 이는 손실 함수에 포함되어 합성 이미지의 인접 픽셀 사이를 비교하여 비교하는 픽셀 사이의 밝기를 낮추는 역할을 통해 합성된 결과물의 품질 향상에 기여할 수 있다. 기존의 종합 변형 정규화를 활용한 연구들에서는 적용 대상 데이터가 사람의 얼굴 사진이기 때문에 바로 인접한 픽셀 사이를 비교하였다. 본 연구에서는 종합 변형 정규화의 적용 대상 데이터가 게임 캐릭터 일러스트 이미지 데이터로 사진에 비해 인접 픽셀 간 색의 차이가 적은 것이 특징이다. 그렇기 때문에 바로 인접한 픽셀 뿐만 아니라 보다 더 넓은 간격의 픽셀과의 비교를 통하여 합성된 결과물의 품질을 보다 더 향상시킬 수 있었다. 실험 결과 최대 5만큼의 픽셀에 대해서 가장 좋은 결과물 합성할 수 있기 때문에 하나의 픽셀과 최대 5까지의 거리만큼 떨어진 픽셀을 비교하여 종합 변형 정규화를 적용하였다.

$$L_{tv} = \sum_{c=1}^{C=3}\sum_{w,h=1}^{W,H}\sum_{k=1}^{K=5} |\hat{I}^f_{w+k,h,c} - \hat{I}^f_{w,h,c}| + |\hat{I}^f_{w,h+k,c} - \hat{I}^f_{w,h,c}| \quad (5)$$

이를 수식으로 (5)와 같이 표현할 수 있다. 이를 통해 합성 이미지에서 발생할 수 있는 인공적인 요소들을 제거하여 보다 더 나은 품질의 이미지를 합성할 수 있게 된다.

$$L_{total} = L_{GAN} + \lambda_1 L_F + \lambda_2 L_{vgg} + \lambda_3 L_{tv} \quad (6)$$

실제로 모델에는 제안된 4가지 손실 함수에 실험을 통해 얻어낸 적절한 가중치를 적용하여, 최종적으로 수식 (6)과 같은 형태의 손실 함수를 사용하였다.

4. 구현 및 결과

4.1 구현 환경

작성된 모델은 파이썬(Python)을 위한 오픈 소스 머신 러닝 라이브러리인 파이토치(PyTorch)[16]를 사용하여 작성되었다. 작성된 모델의 학습과 평가는 4.20GHz Intel(R) Core(TM) i7-7700K CPU, 64GB의 RAM, 2개의 NVIDIA Titan Xp GPU를 탑재하고 있는 우분투(Ubuntu) 운영체제의 컴퓨터에서 실행되었다.

4.2 데이터 셋 정보

작성된 모델은 CASIA-WebFace[17]와 MS-Celeb-1M[18] 데이터 셋으로 학습된 모델을 기반으로 수작업으로 구성한 게임 데이터 셋을 추가로 활용하여 전이 학습시켰다. 수작업으로 구성한 게임 데이터 셋에 포함된 데이터들은 게임 종류에 상관없이 게임 캐릭터의 얼굴 요소(눈, 코, 입, 귀)가 많이 가려지지 않으며 실제 사람의 얼굴 구조와 유사한 캐릭터의 얼굴 부분을 잘라내는 방식으로 구성하였다. 게임 캐릭터 데이터 셋은 작성된 모델이 게임 캐릭터의 질감을 그대로 유지하면서 합성 결과물을 만들기 위해 사용하였다.

4.3 이미지 합성 결과

4.2절에서는 학습된 모델이 실제 게임 캐릭터가 입력으로 주어졌을 때의 합성된 결과가 가정한 것과 같이 잘 동작하는지를 보이기 위해 진행한 모델 평가 결과를 제시하였다.

[Table 2] Experimental results

Input	-30°	-15°	0°	15°	30°

실험을 진행한 데이터의 경우 게임에 상관없이 캐릭터의 눈, 코, 입의 형태와 위치가 명확한 데이터를 선정하였으며, 선정된 게임 캐릭터의 등장 게임 정보에 대해서는 [15]에서 확인할 수 있다. 모델에 입력되는 데이터의 입력 크기는 256 * 256 사이즈, 3채널로 구성되어 있으며 게임 캐릭터의 어깨 위쪽 얼굴의 전체적인 특징을 확인할 수 있는 부분을 잘라(Crop) 사용하였다.

[Table 2]은 학습된 모델에 입력한 데이터에 대한 -30° ~ +30°까지 z축(yaw)으로 회전시킨 결과물을 보여준다. 표를 통해 결과물을 확인할 경우 입력된 이미지의 그림체를 유지하면서도 입력 이미지가 갖고 있는 얼굴의 특징을 잘 보존하여 얼굴을

회전시킨 결과를 합성하는 것을 확인할 수 있다. 실험 결과를 통해 하나의 일러스트를 제안된 모델에 입력하는 것으로 다른 각도의 일러스트 결과를 합성할 수 있는 것을 확인할 수 있다.

5. 결과 분석

본 논문에서 게임 캐릭터의 안면 회전을 하기 위해 제안한 방법이 기존에 존재하는 다른 안면 회전을 위한 방법들에 비해서 어떠한 장점을 갖는지 확인하기 위한 실험을 추가로 진행하였다.

[Table 3] Comparison with other models

Input	DR-GAN	FNM	Ours

이전에 연구되었던 안면 회전 기술 연구 중 실행 가능한 코드를 구할 수 있었던 FNM과 DR-GAN 모델에 대해 본 논문에서 제안한 모델과 동일한 데이터로 실험을 진행하였다. 진행한 실험 결과는 [Table 3]를 통해 확인할 수 있다.

FNM과 DR-GAN의 경우에는 제안 모델과 다르게 정면 얼굴만을 합성할 수 있기 때문에 제안된 모델의 결과 중 정면(0°) 결과와 비교하였다. [Table 2]를 통해 기존의 안면 회전 기술과 제안된 모델 간의 합성된 결과물의 차이를 확인할 수 있다. 제안된 모델에서는 게임 캐릭터의 특징뿐만 아니라 일러스트의 그림체까지 학습하여 결과물을 합성한다. 하지만 이와 달리 FNM과 DR-GAN에서는 입력 이미지가 갖고 있는 신체 부위의 특징을 일부 학습하여 합성된 결과물에 반영하지만 일러스트의 형태가 아닌 사람의 형태의 결과물로 합성하는 것을 확인할 수 있다. DR-GAN의 경우 많은 인공적인 요소를 합성된 결과물에서 확인할 수 있다. FNM의 경우 DR-GAN에 비해서 인공적인 요소들이 포함되어 있는 것은 적으며 입력 이미지가 갖고 있는 얼굴의 특징을 합성된 결과물에 더 잘 반영하였지만 게임 캐릭터의 형태가 아닌 사람의 형태로 결과물을 합성한 것을 확인할 수 있다.

제안된 모델에서는 입력 이미지의 대부분의 특징을 담고 있으면서도 그려진 그림 스타일을 유지한 결과물을 합성한 것을 확인할 수 있다

동일한 데이터로 다른 안면 회전 기술 모델들에 대해서 실험을 진행한 결과 본 논문에서 제안한 모델이 다른 모델들에 비해 입력 이미지에서 확인할 수 있는 특징뿐만 아니라 입력 이미지에서 나타나는 그림 스타일을 잘 유지한 결과물을 합성하는 것을 확인할 수 있다.

6. 결과 및 향후 연구 계획

본 논문은 기존의 사람의 얼굴에 대해서만 구현되던 안면 회전 기술을 게임 캐릭터 일러스트에 대

해서도 동작할 수 있도록 한 기술에 대해서 제안하였다. 이를 증명하기 위해 진행한 실험 결과 대부분의 게임 캐릭터 일러스트에 대해서 다양한 각도의 결과 이미지를 좋은 품질로 합성할 수 있다는 것을 확인할 수 있었다. 제안된 모델을 통해 합성된 결과물은 입력으로 활용된 일러스트가 갖고 있는 대부분의 특징을 잘 보존한다. 하지만 모든 특징을 온전히 갖고 있지는 않다. 하지만 아무 것도 없는 상황에서 다른 각도의 일러스트를 생성하는 것 보다 제안된 방법을 통해 합성된 대부분의 특징이 유지된 다른 각도의 일러스트를 참고하여 일러스트를 작성할 수 있어 일러스트 제작에 소모되는 많은 공수를 아낄 수 있다.

진행한 실험 결과에 대해서 대부분의 게임 캐릭터에서 좋은 결과를 얻을 수 있었지만 코가 점의 형태로 표현되거나 사람이 아닌 게임 캐릭터에 대해서는 좋지 못한 결과를 얻어낼 수 있다는 것을 확인할 수 있었다. 이를 극복하기 위해 결과가 합성되지 못하는 캐릭터 데이터 셋을 포함하여 제안한 모델의 전이 학습(Transfer training)을 진행하여 다양한 사람 형태의 게임 캐릭터에 대해서도 적용 가능한 모델을 학습시키는 방향으로 연구를 진행할 계획이다. 더 나아가 사람의 형태를 띄지 않는 게임 캐릭터들에 대해서도 동작할 수 있는 모델을 학습시킬 수 있는 방법과 실제 유효한 결과를 만들어 낼 수 있는 모델에 대해서도 연구를 진행할 계획이다.

REFERENCES

[1] I.Goodfellow, J.Pouget-Abadie, M.Mirza, B.Xu, D.Warde-Farley, S.Ozair, A.Courville, Y.Bengio. “Generative adversarial networks. ”. Advances in Neural Information Processing Systems 27, 2014.

[2] T.Hassner, S.Harel, E.Paz, R.Enbar. “Effective face frontalization in unconstrained images.” Proceedings of the IEEE conference on computer vision and pattern recognition, pp4295-4303, 2015.

[3] Z.Xiangyu, L.Zhen, Y.Junjie, Y.Dong, L.Stan Z. “High-fidelity pose and expression normalization for face recognition in the wild.” Proceedings of the IEEE conference on computer vision and pattern recognition, pp787-796, 2015.

[4] J.Moniz, C.Beckham, S.Rajotte, S.Homari, C.Pal. “Unsupervised depth estimation, 3d face rotation and replacement.” arXiv preprint arXiv:1803.09202, 2018.

[5] T.Luan, Y.Xi, L.Xiaoming. “Disentangled represerntation learning gan for pose-invariant face recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition, pp1415-1424, 2017

[6] H.Rui, Z.Shu, L.Tianyu, H.Ran. “Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis.” Proceedings of the IEEE international conference on computer vision, pp2439-2448, 2017

[7] H.Yibo, W.Xiang, Y.Bing, H.Ran, S.Zhenan. “Pose-guided photorealistic face rotation.” Proceedings of the IEEE conference on computer vision and pattern recognition, pp8398-8406, 2018

[8] Q.Yichen, D.Weihong, H.Jiani. “Unsupervised face normalization with extreme pose and expression in the wild.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp9851-9858, 2019.

[9] Y.Xi, Y.Xiang, S.Kihyuk, L.Xiaoming, C.Manmohan. “Towards large-pose face frontalization in the wild.” Proceedings of the IEEE international conference on computer vision, pp3990-3999, 2017

[10] D.Jiankang, C.Shiyang, X.Niannan, Z.Yuxiang, Z.Stefanos. “Uv-gan: Adversarial facial uv map completion for pose-invariant face recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition, pp7093-7102, 2018.

[11] C.Jie, H.Yibo, Z.Hongwen, H.Ran, S.Zheman. “Learning a high fidelity pose invariant model for high-resolution face frontalization.” arXiv preprint arXiv:1806.08472, 2018.

[12] Z.Hang, L.Jihao, L.Ziwei, L.Yu, W.Xiaogang. “Rotate-and-render: Unsupervised

photo-realistic face rotation from single-view images." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp5911-5920, 2020.
[13] Z.Jun-Yan, P.Taesung, I.Phillip, E.Alexei A. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE International conference on computer vision, pp2223-2232, 2017.
[14] J.Guo, X.Zhu, Z.Lei. "3DDFA". https://github.com/cleardusk/3DDFA, 2018
[15] Guardian tales, Kong studois, 2020.
Phoenix Wright: Ace Attorney, CAPCOM, 2001.
The war of Genesis, Softmax, 1995.
Hearthstone, Blizzard, 2014.
[16] PyTorch, 2016, https://pytorch.org/
[17] Yi, D., Lei, Z., Liao, S., Li, S. Z. "Learning face representation from scratch." arXiv preprint arXiv:1411.7923, 2014.
[18] Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J. "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition." In European conference on computer vision. pp87-102. 2016.

김 한 동 (Kim, Handong)

약 력 : 2014-2020 상명대학교 컴퓨터과학과 졸업
2020-현재 상명대학교 일반대학원 컴퓨터과학과 석사과정

관심분야 : 소프트웨어 공학, 컴퓨터 비전

한 종 대 (Han, Jongdae)

약 력 : 2005 - 2013 서울대학교 컴퓨터공학과 박사과정
2014 - 현재 상명대학교 컴퓨터과학과 교수

관심분야 : 소프트웨어 공학, 인공지능

양 희 경 (Yang, Heekyung)

약 력 : 2012-2019: 상명대학교 컴퓨터과학과 박사
2019-2020: 상명대학교 컴퓨터과학과 박사후 연구원
2020 - 현재: 상명대학교 SW융합학부 교수

관심분야 : 딥러닝, 컴퓨터 그래픽스, 컴퓨터 비전

민 경 하 (Min, Kyungha)

약 력 : ~ 2000: 포항공과대학교 컴퓨터공학과 박사
2000 ~ 2001: 이화여자대학교 연구교수
2001 ~ 2002: 서강대학교 연구교수
2002 ~ 2006: Rutgers Univ. 박사후 연구원
2006 ~ : 상명대학교 컴퓨터과학과 교수

관심분야 : 컴퓨터 그래픽스, 컴퓨터 비전, 딥 러닝