

# 감성 분류에 따른 GAN 기반 맞춤형 이미지 생성

홍서영, 신슬비, 오승민, 강다빈, 임상순\*

성결대학교

redzerotwo02@naver.com, seulbis108@gmail.com, osm189214@gmail.com, db021129@naver.com,  
\*lssgood80@gmail.com

## An Intelligent Image Generation Scheme based on Emotional Classification with GAN

Seoyeoung Hong, Seulbi Shin, Sheungmin Oh, Dabin Kang, Sangsoon Lim\*

Sungkyul University

### 요약

코로나19 팬데믹 상황으로 인한 가상 공간 생활의 필요성이 대두되며, 메타버스 플랫폼의 중요성이 커지고 있다. 최근 AI 기반 가상 이미지 생성 시스템이 발전함과 동시에 사용자가 원하는 이미지를 능동적으로 생성하는 시스템에 대한 수요도 증가하고 있으며, 이를 반영하여 사용자의 감성 분석 결과를 활용하기 위한 연구 또한 활발하게 진행되고 있다. 따라서 본 논문에서는 감성 데이터와 이미지를 기반으로 감성 정보가 결합된 개인화된 아바타 이미지를 생성함으로써 현실성 높은 콘텐츠 제공을 위한 기법을 제안한다. 이를 위해 사용자들의 대화를 LSTM 모델을 사용하여 문구의 긍정과 부정으로 감성 분류를 진행하고, 분류 결과와 이미지를 StyleCLIP 모델을 이용해 해당 감성에 적합한 사용자 이미지를 생성한다. 이후, Toonify 모델을 이용하여 해당 이미지를 애니메이션화 하여 메타버스 환경에 적용하기 적합한 캐릭터의 모습으로 변환한다.

### I. 서론

코로나19 팬데믹 상황으로 인해 가상의 공간에서 일상생활, 게임, 업무, 소통 등의 생활을 돕는 메타버스의 중요성이 대두되고 있다. 메타버스는 가상을 의미하는 메타(Meta)와 세계나 우주를 말하는 유니버스(Universe)의 합성어이다. 메타버스 속 사람들은 캐릭터를 통해 소통하며 이러한 캐릭터는 곧 자신의 정체성을 표현하기 때문에 독창적이고, 창의력 있는 콘텐츠의 확보가 중요하다[1]. 또한 AI 기반의 가상 이미지 생성 시스템의 발전과 함께 자신이 원하는 특징을 가진 맞춤형 이미지 생성에 대한 수요가 증가하고 있다. 따라서, 본 논문에서는 차세대 딥러닝 알고리즘으로 주목을 받는 LSTM(Long-Short Term Memory)과 GAN(적대적 생성 신경망, Generative Adversarial Network) 기반의 StyleGAN[2]을 사용하여 사용자가 입력한 텍스트의 감성에 따른 맞춤형 이미지를 생성하는 시스템을 제안하며, 이를 통해 캐릭터 콘텐츠 창작 부분의 다양성을 증가시키고자 한다.

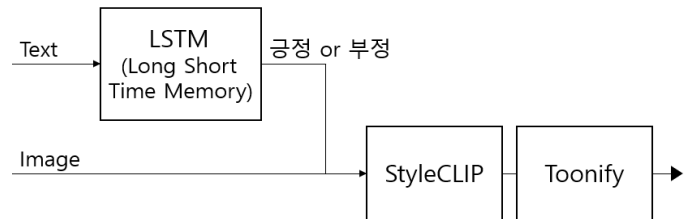
LSTM은 연속적인 시계열 및 자연어 처리를 위한 RNN(Recurrent Neural Networks)의 기울기 소실(Vanishing gradient) 문제를 해결하기 위해 고안된 모델이다. GAN은 노이즈(Noise)를 입력으로 받아 새로운 이미지를 만들어 내거나, 기존의 입력으로 받은 이미지 데이터를 다른 형태의 이미지 데이터로 변환하는 기술을 말한다. GAN은 생성(Generator) 신경망과 판별(Discriminator) 신경망으로 구성된 두 개의 인공 신경망을 적대적으로 학습시켜 실제 데이터와 구분하기 어려운 정도의 정교한 여러 가짜 데이터를 생성한다. StyleGAN은 이미지의 해상도에 따라서 크거나 작은 특징들의 변화를 학습하여, 새로운 이미지를 생성하거나 기존 이미지에 특징을 적용한다.

본 논문에서는 LSTM을 이용한 텍스트 감성 분석 결과와 이미지를 입력으로 받아 GAN을 기반으로 감성 분석 결과에 알맞은 디즈니 풍의 캐리

터 이미지를 추천하는 모델을 제안한다. 해당 기술은 텍스트를 통해 세부적인 감성 표현이 가능하며, 이를 표현하는 독자적인 캐릭터를 생성할 수 있다. 본 연구를 기반으로 메타버스 환경에서 사용자의 대화 도중 감성 표현 여부에 따라 사용자와 유사하며 다양한 감정을 표현할 수 있는 가상 캐릭터 생성을 기대할 수 있다.

### II. 본론

#### 2.1. 시스템 구조

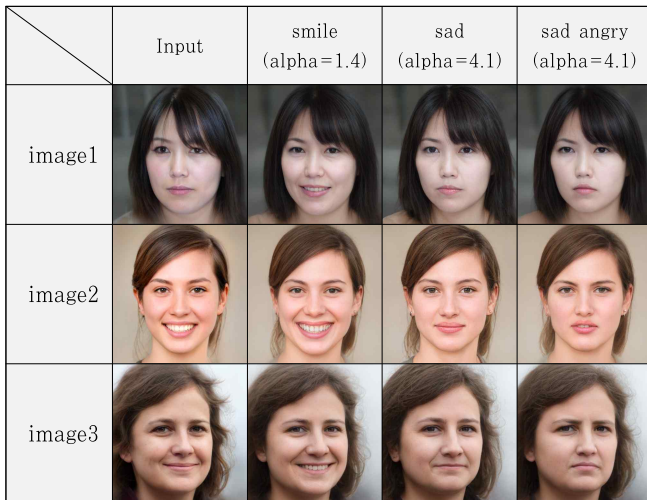


<그림 1. 시스템 구조>

감성 맞춤형 이미지 생성 모델을 위해서는 <그림 1>과 같이 텍스트와 이미지 데이터가 필요하다. 사용자로부터 입력된 텍스트 데이터는 LSTM[3] 모델을 통해 감성 텍스트 데이터로 변환되며, 해당 모델의 경우 긍정과 부정 두 가지로 분류를 진행한다. 변환된 텍스트 데이터는 이미지와 함께 StyleCLIP[4] 모델의 입력으로 사용되고, 해당 출력으로는 감성을 반영한 이미지가 생성된다. 이후 이미지는 Toonify[5] 모델을 통해 디즈니 풍의 이미지로 변환된다.

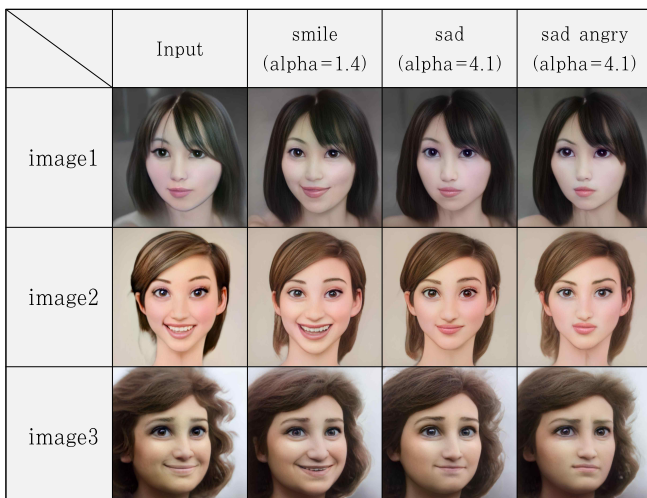
#### 2.2. 연구 내용

본 논문에서 제안하는 시스템은 사용자로부터 텍스트를 입력받아 이를 토큰화 및 분석하여 LSTM 모델과 CNN 모델을 이용해 해당 텍스트의 긍정과 부정을 학습하고, 감성 분류를 진행하며 감성 분류의 결과와 전달받은 이미지를 입력으로 StyleCLIP을 이용해 감성 분류의 결과에 적합한 이미지를 생성한다. 해당 모델은 이미지의 미세조정을 위한 latent vector를 최적화하여 텍스트에서 표현하는 다양한 상황 및 특징을 반영하는 이미지를 생성한다. 이후 이를 기반으로 디즈니 풍의 캐릭터로 변환시키는 Toonify 모델을 사용해 사용자의 감성 정보가 추가된 맞춤 이미지를 생성 및 제공한다.



<그림 2 StyleCLIP 이미지 적용 결과>

StyleCLIP은 StyleGAN과 CLIP을 접목한 모델로, 텍스트 명령어를 이용하여 이미지를 사용자가 원하는 방식으로 조작할 수 있다. <그림 2>는 StyleCLIP을 사용하여 이미지에 텍스트를 적용한 결과이다. smile, sad, angry 세 가지의 감정을 표현할 수 있는 텍스트를 사용하였으며, 이미지 적용 강도의 수치를  $\alpha=1.4$ 와  $\alpha=4.1$ 로 변경하여 진행하였다. 각 입력 이미지에 대한 긍정과 부정 이미지 조작을 통해, 긍정의 경우 치아가 보이며 입꼬리가 올라가는 긍정적인 이미지를 보이며, 부정의 경우 찡그린 표정의 입꼬리가 내려가는 이미지를 보임을 알 수 있다. 따라서 긍정의 경우는 “smiling face”라는 텍스트를 사용하고, 수치는  $\alpha=1.4$ 를 사용하는 것이 적합하며, 부정의 경우 “sad angry face” 텍스트와  $\alpha=4.1$ 의 수치를 사용하는 것이 적합하다.



<그림 3 Toonify 변환 결과>

Toonify는 사람의 얼굴 이미지를 캐릭터로 변환시키는 모델로, 사람 얼굴과 디즈니 캐릭터 이미지를 생성하는 두 개의 StyleGAN 모델을 사용한다. <그림 3>은 StyleCLIP으로 조작된 이미지를 Toonify를 이용하여 디즈니 캐릭터 풍으로 변환시킨 결과이다. 기존의 솔루션을 그대로 사용했을 때 부정적인 표정을 잘 반영하지 못하는 단점이 존재하였고, 이를 보완하기 위해 StyleCLIP을 이용한 이미지 조절 단계에서 두 가지 텍스트를 사용하여 부정적인 이미지를 더욱 극대화하여 적용하였다.

해당 모델은 사용자가 입력한 텍스트의 감성 분류를 진행한 결과와 이미지를 메타버스 캐릭터 생성에 반영할 수 있다. 따라서 자신만의 개성 있는 캐릭터를 생성할 수 있도록 하며, 감성 요소를 반영하기에 더욱 생동감 있는 메타버스 환경을 조성할 수 있다. 다만 기존 모델들의 경우, 이미지 변환과 디즈니가 서양에 초점이 맞춰져 있기에 동양적인 이미지를 입력으로 주었을 경우 캐릭터의 표현 제한적인 경향이 있다. 또한 캐릭터로 변환 시, 모델이 긍정과 부정 두 가지가 아닌 슬픔, 행복, 화남 등의 다양한 표정을 반영할 수 있도록 학습할 필요가 있다. 따라서, 향후 연구에서 텍스트에서의 다중 감성 분류 기법을 적용하여 솔루션을 고도화할 예정이다.

### III. 결론

본 논문에서는 LSTM 모델을 통해 감성 분류를 진행하여 사용자가 입력한 문구를 긍정과 부정으로 분석하고, GAN을 이용하여 맞춤형 이미지를 생성하는 모델을 제안하였다. LSTM, StyleCLIP, Toonify를 활용하여 사용자 맞춤형 새로운 이미지를 생성하는 방안을 모색하였다. 이러한 연구는 메타버스 환경에서 사용자가 자신의 이미지를 기반으로 자신과 유사하며 다양한 감정을 표현할 수 있는 독창적인 캐릭터를 생성하는 방법으로 활용될 수 있다.

논문에서 소개하는 모델은 감성 분류를 긍정과 부정 두 가지로 분류한다는 점에서 한계가 존재한다. 따라서 향후 LSTM이 분류하는 감정의 개수를 늘리고, 그에 맞는 GAN 모델을 설계할 필요가 있다. 또한 모델 생성 시간을 단축하여 이미지 적용의 결과가 빠르게 도출되도록 보완하고자 한다. 추가적으로 해당 모델의 대상을 전신으로 확대하고, 텍스트가 아닌 영상 매체의 음성을 이용한 감성 분류를 진행하여 이미지를 각 감성과 상황에 맞는 캐릭터로 변환시키는 시스템으로 확장할 수 있을 것이다.

### 참 고 문 헌

- [1] 황세연 and 고정훈. (2022). 메타버스환경에서의 캐릭터산업 방향성 연구. 상품문화디자인학연구, 68, 59-73.
- [2] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8110-8119).
- [3] 강다빈. 오승민. 홍서영. 신슬비. (2022). 이커머스 구매전환율 향상을 위한 지능형 고객 관리 시스템. 한국디지털콘텐츠학회. 하계종합학술대회
- [4] Patashnik, O., Wu, Z., Shechtman, E., Cohen-Or, D., & Lischinski, D. (2021). Styleclip: Text-driven manipulation of stylegan imagery. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 2085-2094).
- [5] Justinpinkney. (2020). Github. Toonify .<https://github.com/justinpinkney/toonify>