특집논문 (Special Paper)
방송공학회논문지 제26권 제5호, 2021년 9월 (JBE Vol.26, No.5, September 2021)
https://doi.org/10.5909/JBE.2021.26.5.489
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

인공지능 기반 문화예술 콘텐츠 창작 기술 분석 및 도구 설계

신 춘 성[a)b)], 정 희 용[c)‡]

Analysis and Design of Arts and Culture Content Creation Tool powered by Artificial Intelligence

Choonsung Shin[a)b)] and Hieyong Jeong[c)‡]

요 약

본 논문은 콘텐츠 창작에 새로운 방법과 다양한 가능성을 제공하고 있는 인공지능 기반의 문화예술 콘텐츠 창작기술을 조사하고 이를 바탕으로 일반인을 위한 직관적인 창작도구를 제안한다. 최근 인공지능을 활용한 창작 기술이 다양하게 제안되고 있지만, 정해진 목적으로 개발 및 서비스되고 있어서 창작과 융합 측면에서 확장성과 활용성에 매우 제한이 많다. 본 논문은 인공지능을 바탕으로 한 다양한 데이터 분석과 처리, 콘텐츠 생성 및 창작 및 시각화 기술동향을 살펴보고, 이를 바탕으로 일반인 창작자를 위한 직관적인 창작도구를 제안한다. 제안된 창작도구는 사용자, 창작환경 및 인공지능 특성을 반영하였으며 문화예술 콘텐츠 창작과정에 문화예술 데이터를 처리 및 변환하면서 인공지능 모델을 접목해 새로운 콘텐츠를 생성하기 위한 요소로 구성된다. 이러한 인공지능을 활용한 창작도구는 방대하고 다양한 문화예술 데이터를 효과적으로 처리 및 구조화하고 다양한 생성 및 창작 모델을 적용하여 창작에 필요한 시간을 줄이는 동시에 새로운 아이디어를 실험하도록 지원한다. 제안한 문화예술 콘텐츠 창작도구는 창작자들이 인공지능과 관련 기술을 쉽게 다루면서, 주어진 데이터를 다각적으로 이해하면서 새로운 아이디어와 창의성을 발현하기 위한 초석을 제공할 것으로 기대된다.

Abstract

This paper proposes an arts and culture content creation tool powered by artificial intelligence. With the recent advances in technologies including artificial intelligence, there are active research activities on creating art and culture contents. However, it is still difficult and cumbersome for those who are not familiar with programming and artificial intelligence. In order to deal with the content creation with new technologies, we analyze related creation tools, services and technologies that process with raw visual and audio data, generate new media contents and visualize intermediate results. We then extract key requirements for a future creation tool for creators who are not familiar with programming and artificial intelligence. We finally introduce an intuitive and integrated content creation tool for end-users. We hope that this tool will allow creators to intuitively and creatively generate new media arts and culture contents based on not only understanding given data but also adopting new technologies.

Keyword : art and culture content, artificial intelligence, content creation tool, deep learning

Copyright ⓒ 2021 Korean Institute of Broadcast and Media Engineers. All rights reserved.
“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

Ⅰ. 서 론

4차 산업혁명과 디지털 전환의 가속화로 인해 문화예술 콘텐츠 제작과 향유 방식이 변화하고 있다. 여기에 인공지능, 가상증강현실(augmented reality and virtual reality), 빅데이터, 5G, 블록체인(Block chain) 등 기술혁신과 산업융합에 따라 콘텐츠 기획, 창작, 제작 및 향유하는 방식에서의 변화와 혁신이 가속회되고 있다. 특히 인공지능은 산업수요와 정부정책에 따라 급성장하고 있으며, 인공지능을 탑재한 소비자 응용 분야가 주류로 부상하고 있어 경제적 산업적 가치를 창출할 지능형 콘텐츠 산업 확대가 매우 중요해지고 있다. 더 나아가 최근 디지털 전환과 비대면 산업의 확산과 함께 빅데이터와 알고리즘의 처리 속도가 개선됨에 따라 다양한 데이터의 해석을 바탕으로 하는 문화예술 콘텐츠의 창작과 제작에서도 역할이 가능해지고 있다[1~5].

인공지능을 문화예술에 접목하면 문화예술 데이터 이해를 바탕으로 기획, 창작에 이르는 과정에 고도화와 새로운 경험이 부가된 다양한 형태의 콘텐츠 창작이 가능하다. 인공지능은 문화예술 자료 및 결과물에 대한 영상, 형태, 가치 등 데이터를 획득, 분석, 해석 및 아카이빙하기 위해 활용될 뿐만 아니라, 저장된 문화예술 데이터를 바탕으로 스케치, 검색 등 융합적인 입력을 바탕으로 창작자가 의도하는 스토리와 콘텐츠를 구성하고 추천하는 역할로의 활용도 가능하다[22]. 또한 최근에는 인공지능 기반의 주요 생성 및 변환 모델을 바탕으로 문화예술 콘텐츠를 생성 및 변환하기 위한 2D 영상 생성 및 변환, 3D 영상 생성, 음악 생성 및 변환 등으로 활용이 고도화 되고 있다[20-25].

국내외 많은 연구자들이 문화예술 콘텐츠 창작과정에 인공지능을 융합하여 새로운 형태의 창작과 제작을 시도하고 있다. 대표적으로 구글은 회화, 사진, 음악, 무용 등 다양한 예술 장르에 인공지능을 포함한 최신 기술을 융합하여 문화예술에 대한 접근성을 높이고, 적극적인 참여를 이끌어내기 위한 목적의 Art & Culture Experiment를 프로젝트를 진행하고 있다[5]. 이러한 첨단기술과 문화예술과의 융합은 박물관과 시각예술 콘텐츠의 새로운 방향을 제시하고 있는데, 인공지능을 이용한 키워드 및 유사도 검색, 데이터 시각화, 새로운 오브제 생성, 가상현실과 증강현실 융합 프로젝트 등 새로운 콘텐츠 경험을 실험하고 있다. 여기에 GPU (graphic processing unit)가 내장된 고성능 PC 등 컴퓨팅 장치의 고도화는 미디어 특성을 고려한 고품질 변환, 생성 및 합성 모델을 실현하고 있어 문화예술 콘텐츠 창작과 제작에서의 가능성과 확장성을 높이고 있다. 하지만 문화예술 창작에 인공지능기술 접목은 서로 다른 분야의 협력이 필요한 관계로 활용을 위한 진입장벽이 높아 활용범위가 제한적이다. 이로 인해 인공지능 알고리즘이 빠르게 변화하는 상황과 다른 알고리즘과 상호 연계성 측면이 고려되고 있지 않아 창작에 활용이 어려운 실정이다. 또한 다양한 데이터의 해석과 구조화에 대한 방법이 단편적이고, 다양한 인공지능 모델과의 연계가 어려워 일반인들이 쉽게 융합형 콘텐츠를 창작하기 어려운 실정이다.

본 논문은 인공지능의 가치를 문화예술 분야에 접목하여 새로운 경험과 아이디어가 담긴 콘텐츠를 창작하기 위한 인공지능 기반의 문화예술 콘텐츠 창작방법을 분석하고 이를 바탕으로 통합적 관점에서 창작도구를 제안한다. 이를 위해 2장에서는 먼저 문화예술 콘텐츠 창작과정에 인공지능 활용되고 있는 사례와 활용된 알고리즘에 대해 살펴보고 창작 방식을 분석한다. 3장에서는 인공지능과 데이터를 절차적으로 다루고 창작자의 요구를 맞출 수 있는 미래지향적인 창작도구의 방향을 제안하고 설계한다. 4장에서는 인공지능 기반 문화예술 콘텐츠 창작도구 실현을 위한 향후 연구와 함께 결론을 맺는다.

Ⅱ. 인공지능을 활용한 문화예술 콘텐츠 창작도구 분석

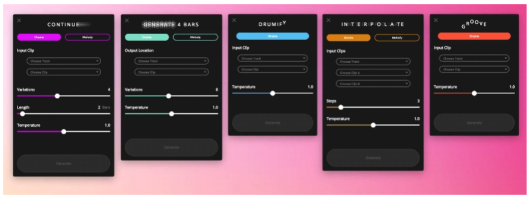
인공지능 기술을 바탕으로 새로운 문화예술 콘텐츠를 생

a) 전남대학교 문화전문대학원, 미디어콘텐츠·컬쳐테크전공 (Program of Media Content and Culture Technology, Graduate School of Culture, Chonnam National University)
b) 전남대학교 일반대학원, 아트앤디자인테크놀로지협동과정 (Interdisciplinary Program of Art & Design Technology, Chonnam National University)
c) 전남대학교 인공지능학부(Department of Artificial Intelligence Convergence, Chonnam National University)
‡ Corresponding Author : 정희용(Hieyong Jeong)
E-mail: h.jeong@jnu.ac.kr
Tel: +82-62-530-3427
ORCID: http://orcid.org/0000-0002-8135-8252
※ This work was financially supported by Chonnam National University (Grant No. 2020-2020).
· Manuscript received July 1, 2021; Revised July 28, 2021; Accepted July 28, 2021.

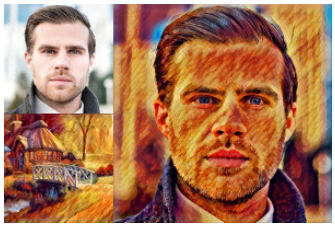
성 및 창작하기 위해 다양한 연구들이 진행되고 있다. 먼저 2016년에 마이크로소프트, 델프트과기대 및 렘브란트 미술관은 렘브란트가 그린 수많은 작품에 대해 얼굴, 자세 등 다양한 특징을 학습하여 렘브란트풍의 영상을 생성하고 3D로 출력하는 넥스트 렘브란트 프로젝트를 진행하였다[1]. 2017년에 구글의 딥마인드팀은 천여 가지의 악기와 30만 가지의 음이 담긴 데이터베이스를 바탕으로 새로운 소리와 음악을 생성하는 마젠타 프로젝트를 수행하였다[2]. 또한 구글은 영상을 바탕으로 딥드림을 생성하는 딥드림 제네레이터(Deep dream generator) 서비스를 시작하였는데, 이 서비스는 사용자가 업로드한 영상에 대해 스타일 영상을 적용하면 스타일이 합성된 영상을 생성하는 서비스와 입력된 영상에 스타일 변환 모델이 적용되어 몽환적인 텍스처가 입혀진 영상을 생성하는 서비스를 제공한다[3]. 최근에 구글은 사용자들이 인터넷 환경에서 직접 인공지능 학습모델을 생성할 수 있는 티처블 머신(Teachable machine)을 개발하였다[4]. 이 서비스는 사용자들이 영상, 사운드, 동작 등의 데이터를 사용자들이 직접 획득하고 이를 바탕으로 인식하는 모델을 학습 및 생성할 수 있는 환경을 제공한다.

선행연구를 살펴본 결과 인공지능을 바탕으로 문화예술 콘텐츠를 이해하고 생성하기 위해 다양한 도구 개발이 진행되었지만 여전히 문화예술 영역에서의 인공지능 활용은 시작단계이다. 먼저 데이터 아카이빙 단계를 보면 저장과 정해진 목적으로의 활용이 우선되고 있어 다양한 형태로의 해석과 활용이 어려운 상황이다. 여기에 저장된 데이터의 검색과 시각화는 기존에 많이 해왔던 영상 및 키워드 방향에서 일부 기능이 추가되고 있다. 검색단계를 살펴보면 인공지능이 결합되어 객체 중심의 검색까지는 왔지만, 다양한 분류나 의미 간 연결된 시각화가 부족하고, 생성과 창작 단계를 살펴보면 새로운 알고리즘을 정해진 데이터에 적용 및 테스트하면서 알고리즘의 성능을 강조하는데 집중하고 있다. 이로 인해 데이터에 담긴 의미와 가치들이 문화예술 콘텐츠 창작에 연계가 어려우며, 콘텐츠 창작과정에 창작자의 해석과 생각을 담아내기 어렵다. 또한 기존 창작자들에게 새로운 시도에 대한 실험과 함께 일반 창작들에게 아이디어를 실현할 기회를 제공하기에는 한계가 있다. 따라서 인공지능 기술을 문화예술 콘텐츠 창작에 효과적으로 융합하는 것이 중요한데, 기획·아카이빙 단계에서 데이터를 구조화하고 해석해서 넣을 수 있는 방안과, 이를 활용해 콘텐츠를 생성 및 창작할 수 있는 도구가 필요한 실정이다.

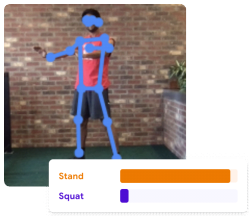
(a)Next rembrandt



(b)Magenta studio



(c)Deep dream generator



(d)Teachable machine

그림 1. 인공지능 기반 문화예술 콘텐츠 창작 도구 (a) 넥스트 렘브란트, (b) 마젠타 스튜디오, (c) 딥 드림 생성기, (d) 티쳐블 머신
Fig. 1. A&C Content Creation based on Artificial Intelligence (a) Next rembrandt, (b) Magenta studio, (c) Deep dream generator, (d) Teachable machine

Ⅲ. 인공지능을 활용한 문화예술 콘텐츠 창작 관련 요소기술 분석

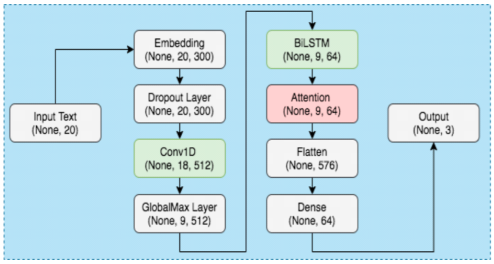
1. 문화예술 데이터 처리

문화예술 콘텐츠 창작을 위해서는 무엇보다도 데이터 확보 및 처리가 매우 중요한데, 이를 위해 시청각 요소를 포함한 데이터를 처리하기 위해 다양한 연구가 진행되고 있다. 문화예술 관련 데이터는 텍스트, 영상 및 음성 등 다양한 형태로 획득이 되고 있어, 이들 데이터 특색에 맞는 처리기법이 연구 및 적용되고 있다. 먼저 인터넷, 서적 및 자료 등에서 가장 많이 발생하는 텍스트는 많은 단어로 구성되어 있어 구조화를 위한 전처리가 필요하다. 이를 위해 언어모델을 바탕으로 n개의 연속된 단어의 빈도를 세는 n-gram 방법, 텍스트 내에 주요단어의 빈도를 파악하는 TF-IDF (term frequency – inverse document frequency), 텍스트에 포함된 내용을 추상화하기 위한 LDA(latent dirichlet allocation) 등의 텍스트 마이닝 알고리즘을 적용되고, 이러한 결과는 텍스트를 요약 및 표현하는 워드벡터 등으로 변환된다[6]. 변환된 처리결과는 고차원의 벡터 결과로 이어지는데 텍스트 이해를 돕기 위해 추가적으로 차원축소를 위한 PCA(principal component analysis) 나 T-SNE(t-distributed stochastic neighbor embedding)를 활용하여 시각화하고 반복된 처리과정을 통해 데이터를 구조화한다[9,28]. 최근에는 딥러닝이 적용되고 있어 텍스트 처리와 모델링의 고도화뿐만 아니라 감성적 분석 등이 시도되고 있다[7-8].

문화예술 관련 데이터는 텍스트뿐만 아니라 비주얼 아트, 조형예술 등 영상 관련 정보들이 많은데, 딥러닝의 발전으로 인해 관련 영상처리에 눈부신 변화가 나타나고 있다. 컴퓨팅 하드웨어 및 신경망 학습 기법의 향상과 함께 심층 신경망(DNN: Deep Neural Networks)의 학습이 가능해지면서, 영상처리에서는 딥러닝을 적용하여 획기적인 발전이 있었다. 특히 2014년 ILSVRC 대회에서 우승한 구글의 Inception 모델과 준우승한 옥스포드 대학의 VGG16 모델을 시작으로 사물인식의 성능이 현저히 높아졌으며[11], 152개의 많은 레이어로 구성된 RasNet이 제안되면서 심지어 인간보다 높은 정확도를 달성하였다[12]. 더 나아가 YOLO2로 인해 사물인식의 종류와 속도가 급격히 향상되어 9000종류의 사물을 실시간으로 구분할 수 있게 되었으며, YOLO3에서는 50fps 수준의 빠른 사물인식을 달성하였다[13-14]. 여기에 사물 및 사람에 대한 3차원 자세인식이 가능해 대상물의 식별뿐만 아니라 공간적 위치와 행동에 대한 이해도가 향상되었다[15]. 이러한 영상처리에 대한 딥러닝 기술은 사물 및 사람에 대한 인식·분류와 3차원 공간에서의 자세에 대한 파악을 가능하게 하여 입력 및 수집된 데이터에 대한 정보를 상세하게 이해하고 구조화하는데 도움을 주고 있다.

영상처리뿐만 아니라 인공지능을 활용해 사운드 데이터를 처리하는 연구도 증가하고 있다. 사운드 데이터는 음성 및 사운드 형태로 mp3, wave 등 저차원에서 만들어진 오디오 파일을 처리하거나 전자악기 표준 포맷인 MIDI(musical instrument digital interface)를 바탕으로 한다. 저차원 오디

(a) LDA



(b) Text modeling based on Deep learning

그림 2. 텍스트 분석 및 처리 기술 (a) LDA 방법, (b) 딥러닝 기반 텍스트 모델링
Fig. 2. Text analysis and processing technologies (a) LDA, (b) text modeling based on deep learning

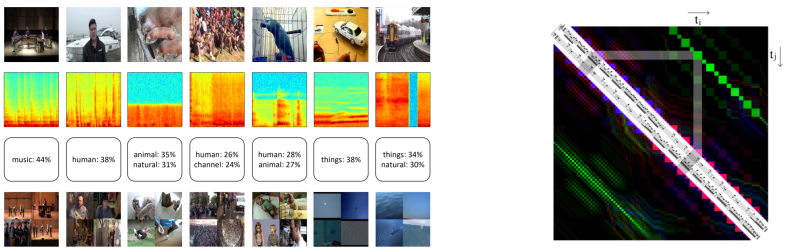
(a) YOLO2 (b) OpenPose

그림 3. 영상 분석 및 처리 기술 (a) YOLO2, (b) OpenPose
Fig. 3. Image analysis and processing technologies (a) YOLO2, (b) OpenPose

오 파일은 그대로 이해 및 활용이 어려워 전처리가 이루어지는데, 시간 도메인 특징, 주파수 도메인 특징, 중심 주파수 성분 등을 추출하여 오디오 스펙트로그램(spectrogram)을 생성한다. 이 스펙트로그램은 사운드가 포함하고 있는 음악적 장르, 악기, 사물의 소리 등의 특징을 내포하고 있어 딥러닝을 적용하여 사운드 기반의 특징 및 이벤트를 식별하는데 활용된다[16]. Tatusuya는 FFT(fast fourier transform)을 통해 주파수 성분으로 구성된 스펙트로그램을 생성한 뒤 인터렉티브 사운드 인식을 위한 시각화로 활용하였다[17]. 최근에는 파이썬 개발환경에 오디오 파일을 처리하기 위한 라이브러리인 liblosa가 제공되고 있어 저차원의 오디오 파일 처리가 한층 더 편리해졌다. 반면 MIDI는 이미 전자적으로 구조화된 악보이며, 곡에 포함된 악기와 리듬 등의 정보를 바로 딥러닝에 적용이 가능하다. 서로 다른 MIDI 파일일지라도 딥러닝을 적용하면 이들 간의 음악적 유사도 및 차이점을 파악할 수 있다[18].

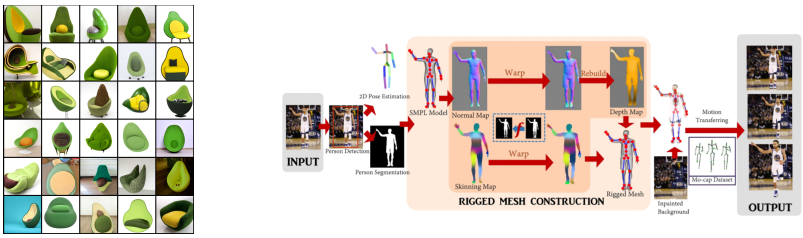
2. 시청각 콘텐츠 생성 및 변환 기술

콘텐츠 생성에서도 가장 주목을 받고 있고 활발한 연구가 진행되는 부분은 인공지능을 활용한 영상 변환 및 생성이다. 구글은 주어진 영상에 사물을 인식하는 모델인 Inception을 적용하여 몽환적인 그림으로 변환 및 생성하는 알고리즘을 개발하였고, 이를 바탕으로 딥드림 서비스를 시작하였다[3]. 튀빙겐대학에서는 입력영상에 스타일 영상을 덧씌워 영상을 생성하는 Neural transfer 알고리즘을 제안하였다[20]. 또한 입력 영상을 바탕으로 출력하는 연구가 진행되고 있는데 구글은 간단한 스케치를 하면 자동으로 관련 영상을 자동으로 생성 및 드로잉하는 Autodraw와 입력된 스케치를 바탕으로 이미지를 자동으로 고화질 영상으로 변환해주는 Pixel2Pixel 알고리즘을 제안하였다[22-23]. 최근에는 2D 정보를 바탕으로 3D를 생성하는 연구가 진행되고 있는데, OpenAI는 텍스트 입력을 이해하고 생성모델을



(a) FFT analysis (b) MIDI comparison

그림 4. 오디오/사운드 분석 및 처리 기술 (a) FFT 분석, (b) MIDI간 비교
Fig. 4. Audio/sound analysis and processing technologies (a) FFT analysis, (b) MIDI comparison



(a) DALL-E (b) Photo Wakeup

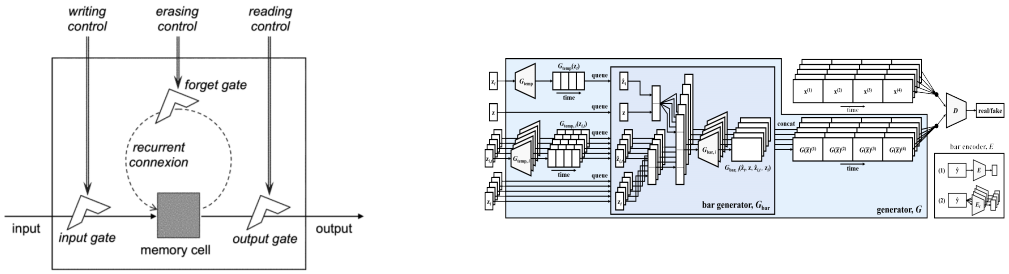
그림 5. 2D/3D 변환 및 생성 기술 (a) DALL-E, (b) Photo Wakeup
Fig. 5. 2D/3D transformation and generation technologies (a) DALL-E, (b) Photo Wakeup

바탕으로 3D 영상을 생성하는 DALL-E 시스템을 제안하였다[24]. 이 시스템은 GPT-3를 바탕으로 텍스트를 이해하고 Text-to-image generation모델을 적용하여 관련된 다양한 3D 형상을 자동을 생성 및 추천한다. 이와 함께 워싱턴대학은 한 장의 사진으로부터 대상물을 인식하고 관련 애니메이션이 가능한 3D 모델을 자동으로 생성 및 증강하는 photo wakeup 기술을 개발하였다[25].

영상뿐만 아니라 사운드 생성 측면에서는 멜로디를 생성하기 위해 순환신경망인 RNN(recurrent neural network)과 적대적 생성모델인 GAN(generative neural network) 등의 딥러닝 모델이 제안되고 있다. RNN은 순환신경망은 입력에 대한 다음을 예측할 수 있어 음성인식, 텍스트인식 등에 많이 활용되고 있는데, 그중에서 학습이 효과적인 LSTM long short-term memory)을 활용한 모델이 주로 이용되고 있다[26]. LSTM 기반의 멜로디 생성 알고리즘은 대표적인 RNN 모델에서 학습효과를 높이기 위해 기억기능을 보강한 모델에 어텐션(attention) 기능을 추가하여 구성된 형태이다. 이를 바탕으로 음악시퀀스를 학습하고, 다음 멜로디를 예측하고 생성한다. 최근에는 적대적 생성모델을 바탕으로 하는 음악생성 모델이 연구되었다. MuseGAN은 화음, 멜로디, 스타일 및 리듬을 입력으로 하는 생성자를 통해 MIDI 마디를 생성하고, 비평자를 통해 원래의 멜로디와 차이가 있는지를 구별하여 생성하는 방식이다[27].

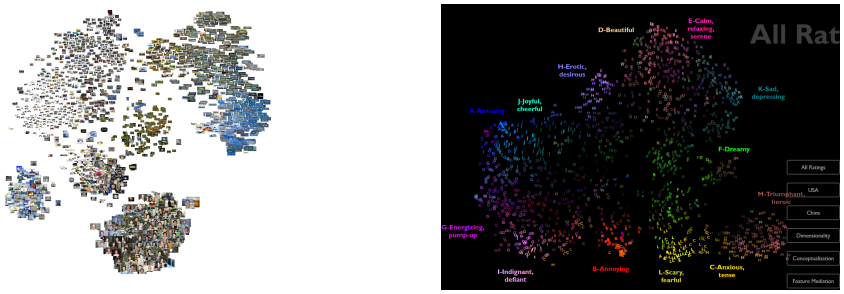
3. 데이터 시각화 기술

창작 과정에서 데이터 문화예술 데이터를 이해하기 위해서는 시각화가 매우 중요한데 다양한 시각화가 진행되고 있다. 기본적으로 T-SNE는 고차원의 데이터를 이해하기 쉬운 2~3차원으로 축소하고 시각화 요소를 조정하면서 다양한 측면으로 데이터를 이해할 수 있도록 하고 있다. 이러한 방법은 텍스트, 영상, 오디오 등 다양한 데이터에 적용되어 이해를 높이는데 활용되고 있다. Karpathy는 트위터 계정에 있는 사용자의 트윗을 획득하여 LDA를 적용한 뒤



(a) LSTM (b) MuseGAN

그림 6. 사운드 및 음악 생성 기술 (a) LSTM 기반 음악생성, (b) GAN 기반 음악 생성
Fig. 6. Sound and music generation models (a) LSTM, (b) MuseGAN



(a) Data visualization based on T-SNE (b) Music Visualization

그림 7. 데이터 시각화 모델 (a) T-SNE기반 데이터 시각화, (b) 음악 데이터 시각화
Fig. 7. Data visualization models (a) Data visualization based on T-SNE, (b) Music mood visualization

T-SNE를 활용하여 사용자간의 관계를 시각화하였다[9]. 텍스트뿐만 아니라 ImageNet이나 MNIST 등의 오픈데이터를 처리하여 특성을 이해하기 위한 시각화가 진행되고 있다[31]. 또한 청취자 입장에서 주관적인 감성적인 분석으로도 접근이 가능한데, 음악전문가들이 미국과 중국관련 사운드 샘플을 청취하고 나서 13개의 문화적·감성적 차이를 정성적으로 평가하여 시각화하였다[19].

문화예술 콘텐츠를 창작하기 위한 데이터 처리, 생성 및 시각화의 과정을 살펴본 결과 데이터를 처리하여 생성하고 시각화하는 과정이 서로 연계되어 있으며, 시청각 데이터들이 복합적으로 활용된다. 또한 이러한 창작의 과정과 함께 창작자들이 쉽게 활용할 수 있도록 하는 사용자 인터페이스와 통합된 저장소가 연계된다.

Ⅳ. 인공지능 기반 문화예술 콘텐츠 창작도구 설계방향

1. 문화예술콘텐츠 창작지원을 위한 요구사항

인공지능을 접목해 문화예술 콘텐츠 창작을 돕기 위해서는 창작자의 특성, 인공지능 특성 및 창작과정에 대한 이해를 바탕으로한 요구사항 정립이 필요하다. 인공지능 측면에서는 다양한 생성, 변환 및 처리 등의 알고리즘을 적절하게 활용할 수 있도록 인터페이스 및 데이터들에 대한 공통 포맷이 필요하다. 또한 인공지능은 학습모델 구성에 시간이 필요하므로 학습된 모델 확보와 이를 적용할 수 있도록 연계가 필요하다.

창작과정 특성 측면에서는 문화예술 데이터 이해, 새로운 아이디어 도출과 실험 및 창작 과정이 반복될 수 있도록 지원이 필요하다. 이를 위해서는 다양한 데이터 입력으로부터의 처리하여 공통 데이터 포맷에 대한 변환 및 저장이 필요하고, 중간 결과물을 저장하고 다시 처리할 수 있도록 저장소 관리(클라우드 저장소)가 필요하다. 그리고 이러한 데이터의 이해, 스토리 구성 및 생성을 꾸준히 반복하는 과정이고, 중간 결과물을 활용해 확장과 조합하여 변경해나갈 수 있도록 창작 파이프라인 정립과 지원이 필요하다.

창작자의 특성 측면에서는 무엇보다도 직관적인 통합 창작 환경이 필요하다. 직관적인 통합 창작 환경은 무엇보다도 사용자 인터페이스가 중요한데, 직관적인 데이터의 이해를 바탕으로 콘텐츠 생성과, 데이터의 지능적 및 다차원 시각화를 지원해야 한다. 그리고 상세하고 정밀한 고수준의 콘텐츠 생성을 위한 코딩환경으로의 연계도 필요하다. 통합창작 환경에서 기본적인 데이터처리 및 생성을 지원한다고 하면, 코딩환경에서는 자유롭고 정밀한 콘텐츠 생성을 지원해야 한다.

2. 인공지능 기반의 문화예술 콘텐츠 창작도구 설계안

인공지능을 기반으로 문화예술 콘텐츠를 창작하기 위한 도구는 인공지능 측면, 창작특성 지원, 직관적인 통합 창작 환경을 지원하기 위해 데이터 입력과 전처리, 데이터 검색 및 시각화, 콘텐츠 생성과 융합, 공통 데이터 및 저장소 및

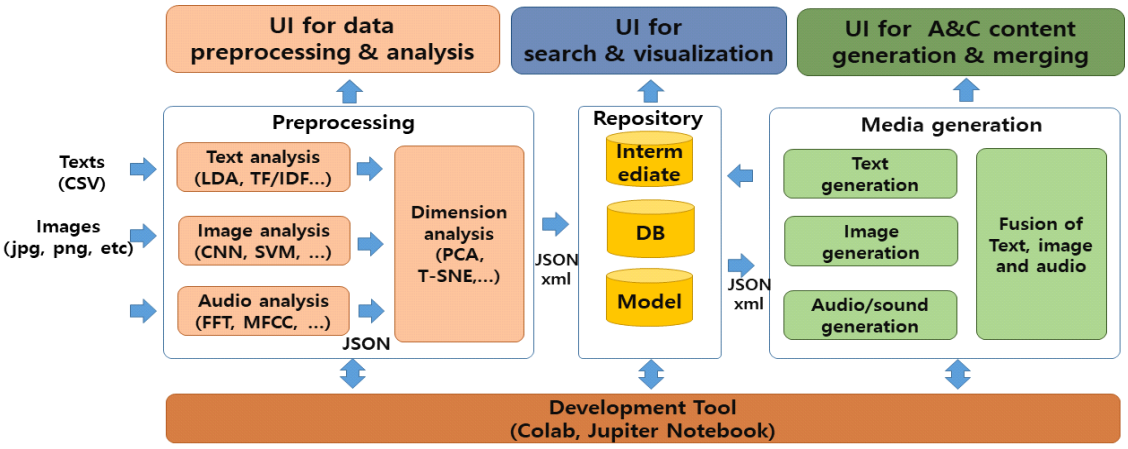


그림 8. 인공지능 기반 문화예술 콘텐츠 창작을 위한 통합형 창작 도구 구조
Fig. 8. Architecture for A&C content creation based on artificial intelligence

통합 창작 환경으로 구성된다. 데이터 입력은 텍스트, 영상, 오디오 등 다양한 데이터에 대한 입력이 가능하고, 전처리는 개별 미디어 특성에 따른 전처리 기능과 함께 데이터 차원을 바탕으로 한 시각화가 수행된다. 텍스트의 경우 CSV(comma separated value) 일반 텍스트 등에 대한 필터링, 변환 등의 전처리와 다차원 시각화 분석을 통해 구조화되어 저장된다. 영상 데이터의 경우 다양한 영상 에 대한 포맷변환과 전처리를 통해 추후 분석 및 시각화를 용이하게 한다. 음성 및 오디오 데이터는 오디오 포맷 변환과 분석이 수행된다.

검색 및 시각화에서는 생성된 결과에 대한 다양한 측면의 이해를 위한 검색과 다차원적인 분석이 진행된다. 주요성분 분석인 PCA나 T-SNE 등의 차원축소를 통해 데이터 이해와 주요성분 분석이 수행된다.

창작단계에서는 이러한 처리 결과를 바탕으로 관련 인공지능 알고리즘을 활용하여 시각, 청각 및 텍스트 요소를 생성하고, 필요에 따라서는 이들 간의 연계와 융합해한 콘텐츠 생성으로 이어지도록 한다. 텍스트의 경우 LSTM, GPT-2(generative pre-trained transformer 2), GPT-3(generative pre-trained transformer 3) 등의 생성 알고리즘을 통해 대상 콘텐츠에서 필요한 새로운 텍스트를 자동으로 생성한다. 영상의 경우 스타일 변환, 스케치에서 영상 생성, 2D에서 3D 생성과 변환 등의 알고리즘이 연계된다. 오디오의 경우 분석된 오디오 데이터를 바탕으로 음악적 요소를 활용하여 새로운 배경음악이나 멜로디 생성이 이루어진다. 또한 영상, 텍스트 및 음성 특성을 연계하여 관련된 미디어를 융합적으로 생성되도록 지원한다.

여기에 데이터를 저장하고 관리하기 위한 저장소는 인공

표 1. 인공지능 기반 미디어 데이터 처리 및 생성을 위한 모델
Table 1. Models for processing and generation of media data

Purpose	Related tools
AI Art	Runway ML, Nature of Code, GANBreeder, Magenta, Processing, ml5.js, etc
Music / Sound	Magenta Studio, AI Duet, NSynth Sound Maker, MuseNet, Pitch Detection, etc
Movement / Dance	AI Sketches With Bill T.Jones, etc
Voice	Scribbling Speech, Thing Translator, etc
Images / Pictures	Deep Dream Generator, DeepArt.io, Visionist, GoArt, Deep Angel, Google Deep Dream, GANBreeder, Fractal Art Generators, etc
Data Visualization	Activation Atlases, What neural networks see, Visualizing high dimensional space, t-SNE, etc
AI Drawing / AI Painting	AutoDraw, AI Painter, Quick-Draw, Sketch-RNN Demos, Cartoonify, etc

지능 처리 및 생성 관련 모델과 데이터를 저장하고, 처리 및 생성을 위한 중간 데이터를 저장한다. 또한 창작과정에서 생성된 중간 데이터를 단계별 및 다른 모델 간에 공유하기 위해서는 공통된 데이터 포맷이 중요한데 JSON(java-script object notation)이나 XML(extensible markup language)행태로 표현 및 저장한다. JSON은 네트워크 환경에서 데이터를 교환하기 위한 포멧으로 시스템 내외부에 활용이 가능하고 XML은 더 풍부한 메타데이터를 제공한다. 창작과 관련된 인공지능 모델은 인터넷과 깃허브(Github)를 통해 확보가 가능하다. 위의 표1은 인터넷과 깃허브를 통해 공개되어 누구나 다 사용할 수 있는 대표적인 모델을 나타내고 있다.

마지막으로 창작과정 지원뿐만 아니라 창작자의 의도를 반영하고 인공지능의 특징을 잘 이하면서 활용을 돕기 위해서는 직관적 인터페이스와 통합적 창작환경이 중요하다. 직관적 인터페이스는 데이터 입출력, 전처리를 위한 알고리즘 선택 및 적용, 미디어 종류에 따른 인공지능 처리 및 생성 모델의 선택 및 처리 등의 다양한 창작요소를 쉽게 다룰 수 있도록 2D/3D 시각화, 그래픽환경 및 플러그인 형태로 구성된다. 또한 상황에 따라 상세한 프로그래밍을 할 수 있도록 클라우드 기반 개발 환경이나 로컬저장소 기반의 개발 환경이 연동된다.

이러한 통합 창작 환경에서 인공지능 기술을 활용하고자 하는 창작자와 일반사용자들은 인공지능 특성과 데이터를 이해하면서 관심 콘텐츠에 대한 아이디어를 발현 및 실현하여 창작해나갈 수 있다. 창작자들은 문화예술 관련 시청각 및 텍스트 데이터를 다양한 관점에서 획득 및 이해하고, 분석과 시각화를 과정을 거치면서 아이디어를 찾고, 다양한 형태의 반복된 창작과 실험이 가능하다. 이를 통해 제안한 창작도구는 융복합 교육, 전시체험, 미디어아트, 아트앤테크놀로지 등의 문화예술 융복합콘텐츠를 실험하고 경험을 확대하는데 역할이 가능할 것으로 기대된다.

V. 결 론

본 논문은 인공지능을 기반으로 문화예술 콘텐츠를 창작하기 위한 요소 기술을 조사하고 이를 바탕으로 창작자의 의도와 경험을 반영하기 위한 통합적 및 직관적 창작도구를 제안하였다. 최근 인공지능을 활용하면서 복잡하고 반복적인 처리를 줄일 수 있고, 방대한 데이터를 바탕으로 새로운 경험과 아이디어가 포함된 콘텐츠의 창작과 제작이 진행되고 있다. 그럼에도 인공지능을 기반으로 한 창작을 위해서는 코딩과 함께 인공지능에 대한 많은 이해를 요구하고 있어 일반인 및 기존의 창작자들의 진입이 어려운 실정이다. 본 논문은 인공지능을 바탕으로 다양한 시청각 문화예술 데이터를 처리하는 기술, 생성 및 변환하는 기술 및 시각화하는 기술을 조사하였으며 이를 바탕으로 코딩과 인공지능에 익숙하지 않은 창작자들이 보다 쉽게 아이디어를 내고 창의성을 향상하여 콘텐츠를 창작할 수 있는 창작도구의 방향을 제안하였다. 제안한 창작도구는 인공지능의 특성, 사용자의 특성 및 창작과정 특성을 반영하여 구성하였으며, 이를 통해 창작자들이 창작관련 데이터와 인공지능 기술을 보다 쉽게 활용하면서 창작의도를 반영할 수 있는 창작환경 제공을 목표로 하였다.

제안한 도구는 이제 시작단계에 있어, 향후 추진할 연구로는 다음과 같다. 먼저 인공지능을 포함한 기술의 발전 및 진화가 빠르게 되고 있어 제안한 도구의 실체화와 현실적인 활용성에 대한 검토가 필요하다. 또한 주요 장르나 기술을 중심으로 아카이빙, 창작 및 시각화를 통합해 실현함으로써 전체 구조와 사용자의 특성을 반영할 수 있도록 해야 한다. 그리고 장르간이나 데이터 요소간 연계를 바탕으로 한 창작을 지원하기 위한 융합적 생성 방법도 연구될 예정이다.

참 고 문 헌 (References)

[1] Google Arts & Culture, https://artsandculture.google.com/ (accessed June. 27, 2021)
[2] The Next Rembrandt, https://thenextrembrandt.pr.co/(accessed June. 27, 2021)
[3] Magenta, https://magenta.tensorflow.org/(accessed June. 27, 2021)
[4] Deep dream generator, https://deepdreamgenerator.com/(accessed June. 27, 2021)
[5] Teachable Machine, https://teachablemachine.withgoogle.com/(accessed June. 27, 2021)
[6] D. M. Blei, A. Y. Ng, M. I. Jordan, “Latent Dirichlet Allocation,”

Journal of Machine Learning Research, Vol. 3, pp. 993-1022, 2003.

[7] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep Learning-based Text Classification: A Comprehensive Review. ACM Computing Survey,” Vol. 54, No. 3, Article 62, 2021. doiI:https://doi.org/10.1145/3439726

[8] Z. Kastrati, L. Ahmedi, A. Kurti, F. Kadriu, D. Murtezaj, F. Gashi, “A Deep Learning Sentiment Analyser for Social Media Comments in Low-Resource Languages,” MDPI Electronics, Vol. 10, No 1133, 2021. https://doi.org/10.3390/electronics10101133

[9] tSNE Java Script demo, https://cs.stanford.edu/people/karpathy/tsnejs/ (accessed June. 27, 2021)

[10] C. Sievert, K. Shirley, “LDAvis: A method for visualizing and interpreting topics,” In Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp. 63-70, 2014.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” In Proc of the 25th International Conference on Neural Information Processing Systems(NIPS'12), Vol. 1,Curran Associates Inc., Red Hook, NY, USA, 1097–1105.

[12] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," In Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016. doi: 10.1109/CVPR.2016.90.

[13] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," In Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517-6525, 2017. doi: 10.1109/CVPR.2017.690.

[14] J. Redmon, A. Farhadi, “YOLOv3: An Incremental Improvement,” ArXiv abs/1804.02767, 2018.

[15] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," In Proc of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 2021. doi: 10.1109/TPAMI.2019.2929257.

[16] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. Weiss, K. Wilson, "CNN architectures for large-scale audio classification," In Proc of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 131-135, doi: 10.1109/ICASSP.2017.7952132.

[17] T. Ishibashi, Y. Nakao, and Y. Sugano, “Investigating audio data visualization for interactive sound recognition,” In Proc of the 25th International Conference on Intelligent User Interfaces (IUI), Association for Computing Machinery, New York, NY, USA, 67–77. doi:https://doi.org/10.1145/3377325.3377483

[18] Z. Wolkowicz, S. Brooks, V. Keselj, “Midivis: Visualizing Music Structure via Similarity Matrices,” In Proc of International computer music conference(ICMC), pp. 53-56, 2009.

[19] A.S. Cowen, X. Fang, D. Sauter, D. Keltner, “What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures,” In Proc of the National Academy of Sciences, Vol. 117(4), pp. 1924-1934, 2020. doi: 10.1073/pnas.1910704117.

[20] L. A. Gatys, A. S. Ecker, M. Bethge, “Image Style Transfer Using Convolutional Neural Networks,” In Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2414-2423, 2016.

[21] P. Isola, J. Zhu, T. Zhou and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” In Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967-5976, 2017. doi: 10.1109/CVPR.2017.632.

[22] Autodraw, https://www.autodraw.com/ (accessed June. 27, 2021)

[23] P.Isola, J. Zhu, T. Zhou, A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” In Proc of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967-5976, 2017. doi: 10.1109/CVPR.2017.632.

[24] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, “Zero-Shot Text-to-Image Generation,” ArXiv, abs/2102.12092. 2020.

[25] C. Weng, B. Curless, I. Kemelmacher, “Photo Wake-Up: 3D Character Animation From a Single Photo,” pp. 5901-5910, 2019. doi: 10.1109/CVPR.2019.00606.

[26] J.P. Briot, G. Hadjeres, F. Pachet, “Deep Learning Techniques for Music Generation - A Survey,” Computational Synthesis and Creative Systems, Springer, 2017.

[27] H. Dong, W. Hsiao, Li Yang, and Y. Yang, “MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment,” In Proc of the 32nd AAAI Conference on Artificial Intelligence (AAAI), 2018.

[28] L. Maaten and G.E. Hinton. Visualizing High-Dimensional Data Using t-SNE, Journal of Machine Learning Research, Vol 9, pp. 2579-2605, 2008.

[29] M. Wattenberg, F. Viegas, I. Johnson, “How to Use t-SNE Effectively,” Distill, 2016. http://doi.org/10.23915/distill.00002 (accessed June. 27, 2021)

[30] T-SNE visualization, https://lvdmaaten.github.io/tsne/ (accessed June. 27, 2021)

[31] Music visualization, https://www.ocf.berkeley.edu/~acowen/music.html# (accessed June. 27, 2021)

저 자 소 개

신 춘 성

- 숭실대학교 컴퓨터학부 학사
- 광주과학기술원 정보통신공학 석사
- 광주과학기술원 정보통신공학 박사
- 카네기멜론대학 HCI Institute 박사후연구원
- 한국전자기술연구원 VR/AR연구센터 책임연구원
- 문화체육관광부/한국콘텐츠진흥원 문화기술 PD
- 현재 : 전남대학교 문화전문대학원 미디어콘텐츠·컬쳐테크전공 부교수
- 현재 : 전남대학교 일반대학원 아트앤테크놀로지협동과정 겸임교수
- 현재 : 전남대학교 문화기술연구소장
- ORCID : http://orcid.org/0000-0003-2384-4022
- 주관심분야 : 가상증강현실, HCI, 문화기술

정 희 용

- 부경대학교 제어계측공학과 학사
- Hiroshima University 로봇공학과 석사
- Osaka University 기계공학과 박사
- 삼성중공업 산업기술연구소 책임연구원
- Osaka University 의학계연구과 교수
- 현재 : 전남대학교 AI융합대학 인공지능학부 교수
- ORCID : http://orcid.org/0000-0002-8135-8252
- 주관심분야 : 헬스케어, 지능로봇, 컴퓨터비전