

## 비대면 소통 플랫폼 맞춤형 영상 합성 기술 연구 : 디지털 문화 소비의 주요 계층인 MZ세대를 중심으로

황 서 빈<sup>1</sup> · 김 혜 근<sup>2</sup> · 윤 경 윤<sup>1</sup> · 황 중 원<sup>2</sup> · 조 영 준<sup>2\*</sup>

<sup>1</sup>전남대학교 IoT인공지능융합학과

<sup>2</sup>전남대학교 소프트웨어공학과

### A Research on Group Photo Synthesizer in Untact Communication Platform : focusing on generation MZ

Seo-Bin Hwang<sup>1</sup> · Hye-Geun Kim<sup>2</sup> · Kyung-Yoon Yoon<sup>1</sup> · Jong-Won Hwang<sup>2</sup> · Yeong-Jun Cho<sup>2\*</sup>

<sup>1</sup>Undergraduate Student, Department of IoT Artificial Intelligence Convergence, Chonnam National University

<sup>2</sup>Undergraduate Student, Department of Software Engineering, Chonnam National University

<sup>2\*</sup>Assistant Professor, Department of Software Engineering, Chonnam National University

#### [요 약]

전례 없는 코로나 팬데믹 상황에서 질병 전파를 늦추기 위한 사회적 거리두기 감염 관리가 확대되고 있으며 이에 따라 비대면 소통 플랫폼의 모든 연령층에 대한 사용량이 급증하였다. 특히 디지털 트렌드를 주도하는 MZ세대는 온라인에서의 다양한 소통과 만남이 일상화되었으나 일반적인 비대면 소통 플랫폼에서는 다양한 문화를 지닌 MZ세대의 요구를 충족시키기 어렵다. 이를 해결하기 위해 본 연구에서는 "비대면 소통 플랫폼 맞춤형 영상 합성 기술"이라는 새로운 형태의 디지털 콘텐츠 생성 기술을 제안한다. 해당 기술에서는 강화된 객체 검출 기법을 통해 사용자의 위치를 탐지하고 최적의 인물 배치 및 합성영상 생성을 모두 자동으로 수행한다. 서비스 사용에 대한 높은 편의성 및 접근성 위함 모든 기술은 웹(Web)에서 활용이 가능하도록 구성됐다. 제안 기법 및 서비스는 다양한 입력 영상 및 환경에 대해 올바르게 단체 영상을 합성하는 것을 확인하였으며 본 연구의 결과물에 대한 사용자들의 다양한 설문문을 통해 연구 필요성 입증 및 연구 결과물의 활용성에 대한 검증을 수행하였다.

#### [Abstract]

To slow the spread of the disease under COVID-19 pandemic, infection control called "social distancing" has expanded. Since then, the use of the communication platforms have soared inevitably. Likewise, various online meeting platforms are commonly provided, but the existing platforms can not satisfy generation MZ and their diverse cultures. In this work, we propose a new service called "Group Photo Synthesizer" that automatically synthesizes a group photo from individual person images. To this end, an improved person detector automatically detects positions of people in images. Next, the individual images will be properly located and synthesized as a group photo. For user convenience, all proposed methods are designed to be available on the web platform. In addition, we extensively validated the proposed methods under various input images and conditions. Furthermore, through various surveys of users on the results of this study, the necessity of research was proved and the usability of the research results was verified.

**색인어** : 비대면 소통 플랫폼, MZ세대, 인공지능, 단체 인물사진 합성 기술, 컴퓨터 비전

**Keyword** : Non-face-to-face Communication Platform, Generation MZ, AI, Group Photo Synthesizing, Computer Vision

<http://dx.doi.org/10.9728/dcs.2021.22.12.2139>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Received** 13 October 2021; **Revised** 10 November 2021

**Accepted** 10 November 2021

**\*Corresponding Author; Yeong-Jun Cho**

**Tel:** +82-62-530-3432

**E-mail:** yj.cho@chonnam.ac.kr

## I. 서론

2020년 3월, 세계보건기구 WHO는 사상 세 번째 팬데믹(Pandemic)을 선포[1]했다. 이는 신종 코로나 바이러스 감염증-19(이하 코로나19)로 인한 감염세가 전 대륙에 확대되어 내려진 결정이었다. 전례 없는 코로나19 팬데믹으로 인해 사람 간의 접촉을 감소시켜 질병의 전파를 늦추는 “사회적 거리 두기(Social distancing)”라는 감염 관리가 전 세계적으로 확대되었다[2]. 이러한 관리 체계는 전염병 예방에는 효과적이나 사람 간 직접적인 소통이 어려워지는 단점이 있다.

따라서 비대면 소통 플랫폼을 사용해 소통하는 사례가 급격히 증가하고 있다. 비대면 소통 플랫폼이란 다수의 사용자가 실시간으로 화상 및 음성 등을 주고받으며 소통을 할 수 있는 플랫폼으로써 줌 비디오 커뮤니케이션 社의 Zoom 및 구글 社의 Meet 등이 대표적인 예라고 할 수 있다. 최근 코로나 팬데믹과 비대면 플랫폼의 확대가 맞물려 디지털 트렌드를 주도하는 MZ세대(Generation MZ)는 온라인에서의 만남이 일상이 되었다. 대학을 중심으로 사용되는 수업과 소모임 활동은 물론, 회사업무, 술자리와 같은 모든 사회 활동이 비대면 소통 플랫폼에서 진행되고 있다.

이렇듯 디지털 환경에 익숙한 MZ세대는 비대면 소통 플랫폼을 이용한 소통에 매우 긍정적[3]이다. 이들은 팬데믹으로 인해 모임을 할 수 없거나 여행을 갈 수 없는 상황에서 이에 대한 갈증을 “랜선 회식” 및 “랜선 여행”이라는 새로운 문화를 통해 해소한다. 랜선이라는 수식어는 온라인을 통해 비대면으로 해당 행동을 하는 것을 의미한다. 예를 들어 랜선 여행은 온라인으로 본인이 방문하고 싶은 명소를 방문하는 것을 뜻한다. 재미를 추구하는 MZ세대는 온라인 소통 플랫폼 내부의 가상공간에서 함께한 자신과 지인들의 모습을 합성시켜 단체 기념사진을 만들고 이를 공유하는 문화를 만들기도 하였다.

앞서 살펴보았던 것처럼 MZ세대는 코로나 팬데믹의 상황에서도 비대면 소통 플랫폼을 창의적으로 활용해 인적 네트워크를 유지한다. 또한, 기념사진 만들기와 같은 다양한 방법으로 자신의 삶을 타인과 공유하고 있으며 이는 하나의 문화로서 자리 잡았다. 그러나 기존의 비대면 소통 플랫폼은 일반적으로 그림 1.(좌)와 같이 사용자가 동일한 규격의 직사각형 칸 안에 한 명씩 정렬된 화면으로 구성된다. 이러한 플랫폼에서 자연스러운 기념사진을 제작하기 위해서는 모든 작업을 제작자가 직접 진행해야 하며 많은 노력과 시간이 소요된다. 이는 즐거움과 속도감을 추구하는 MZ세대의 요구를 충족하지 못한다.

본 연구에서는 이와 같은 문제들을 검토하여 여러 사람의 영상을 자동으로 합성하고 자연스러운 기념사진을 생성해주는 “비대면 소통 플랫폼 맞춤형 영상 합성 기술”을 제안한다(그림 2). 해당 기술은 사용자의 조작을 최소화하여 매우 간편하게 동작한다. 이를 구현하기 위해 다양한 최신 영상 처리 및 영상 이해 기술 등을 활용하였으며 다양한 기기로 자유롭게 활용할 수 있도록 전체 구성요소를 웹(Web)상에 구축하였다.

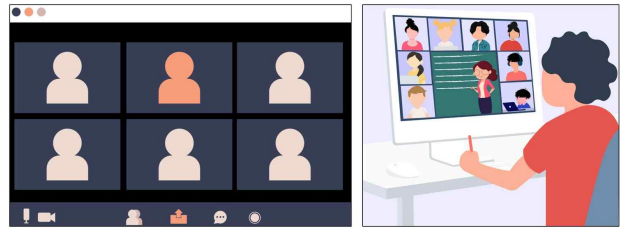


그림 1. (좌) 비대면 소통 플랫폼의 제공 화면 예시  
(우) 서로 다른 독립 공간에서 비대면 소통 플랫폼에 참여

Fig. 1. (Left) Untact meeting platform interface  
(Right) Each participant accesses the platform in an independent space

제한한 서비스에서는 먼저 입력받은 영상들에서 사람의 위치를 객체 검출기[4]를 통해 자동으로 검출한다. 단, 비대면 플랫폼 환경에서는 사람이 상반신만 노출되는 경우가 많으므로 해당 상황에서도 높은 검출 정확도를 수행하도록 검출기를 강화하였다. 그다음 검출된 사람영역은 영상 분할 [5,6] 기술을 통해 배경과 분리가 되며 이와 동시에 사람의 얼굴 크기가 측정된다. 측정된 얼굴 크기와 합성하는 인물의 수를 고려하여 제안하는 기술에서는 올바르게 사람들을 합성할 수 있는 최적의 위치를 추천한다. 최종적으로 서로 다른 공간에서 각각 촬영된 인물들의 영상(그림 1.(우))들이 입력되어도 자연스러운 합성 결과물을 얻기 위해 영상 스타일 변환기술을 [7,8] 적용한다.

이렇듯 본 연구에서는 다양한 최신 인공지능 및 영상이해 기술들을 기반으로 MZ세대의 요구를 충족시킬 수 있는 새로운 디지털 콘텐츠 생성 서비스 줌잇추(ZOOMMEETCHU; 줌에서 만남을 추구)를 제안한다. 동시에 해당 기술을 통해 제공하는 서비스는 웹서비스 형태로 구축하여 다양한 상황 및 플랫폼에 적용 가능한 높은 이식성을 보인다. 본 연구에서 제안하는 기술 및 서비스가 사회적 거리 두기 단계가 심화되고 있는 코로나 팬데믹 상황에서 비대면 소통 플랫폼 사용을 장려하고 많은 사람에게 재미와 활력을 줄 수 있기를 기대한다.

이어지는 본 논문의 제2장에서는 본 연구의 이론적 배경에 대하여 논의한다. 팬데믹 현상을 중심으로 비대면 소통 플랫폼을 살펴본 후, MZ세대의 특징과 문화를 살핀다. 이를 기반으로 기존 비대면 소통 플랫폼 및 인물 합성 서비스의 문제점에 대해 고찰한다. 또한, 본 연구에서 활용된 주요 인공지능 기술들에 대해 논의한다. 제3장에서는 제안하는 기술의 구성과 알고리즘, 그리고 실제로 구현된 서비스에 관해 서술한다. 제4장에서 제안 알고리즘의 실행 결과 및 활용에 대해 살펴본다. 마지막으로 본 연구에서 제안하는 기술의 활용 및 기대효과에 대해 설명하고 차후 연구의 방향성을 제시하며 마무리한다.

## II. 배경 연구

### 2-1 비대면 소통 플랫폼과 MZ세대의 문화

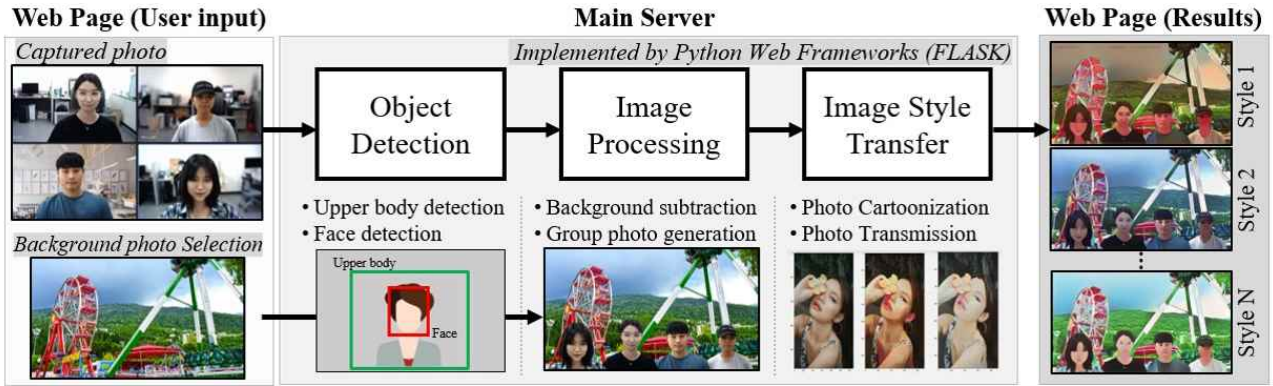


그림 2. 제안하는 비대면 소통 플랫폼 맞춤형 영상 합성 서비스  
 Fig. 2. Proposed framework for automatic group photo synthesizing service

코로나바이러스 감염증-19(이하 코로나19)는 2019년 12월 중국 후베이성 우한시에서 원인불명의 폐렴이 집단으로 발병하면서 시작되었다. 이후 중국 전역은 물론 한국을 포함한 주변 아시아 국가 등으로 감염세가 확산되었다. 결국, 전 대륙에서 감염자가 속출함에 따라 세계보건기구 WHO는 사상 세 번째로 팬데믹을 선포[1]했다. 특히 2021년 7월을 기점으로 대한민국에서 감염자의 수가 급증했으며, 코로나19 감염 우려로 인해 정부는 “사회적 거리두기”[2] 규제를 강화했다. 그로인해 직접 대면하지 않고 사람과 소통할 수 있는 비대면 소통 플랫폼의 사용량이 급격히 증가했다. 코로나 팬데믹 상황 이후 세계보건기구(WHO)는 평균 기온이 1도 올라갈 때마다 감염병 발생률이 4.7% 증가함을 밝혔다[9]. 즉, 지구 온난화가 심화되고 있는 현 상황에서는 코로나 팬데믹이 종료되더라도 다른 감염병 팬데믹이 발생할 가능성이 점차 커지고 있음을 의미한다. 따라서 현재와 같은 감염병 사태가 추후에 발생하더라도 비대면 상황에서 인적 네트워크가 정상적으로 유지될 수 있도록 지속적인 비대면 소통 플랫폼 구축 및 관련 서비스 확대가 필요한 실정이다.

“MZ세대”란 인터넷 문화가 익숙한 “밀레니얼 세대”와 스마트폰과 같은 디지털 문화가 익숙한 “Z세대”가 결합한 용어로 1981~2010년생을 말한다. MZ세대의 취향과 소비 형태는 다양한 분야에 영향을 미치고 있고 이들의 문화적 영향력은 현재 어느 계층보다 가장 중요하게 인식되고 있다[10]. 우리금융경영연구소에는 최근 비즈니스 생태계 트렌드가 “MZ세대 문화”임을 전망했다. MZ세대는 “① 재미를 추구하며 사교가 자유롭고, ② 짧은 호흡과 속도감, 변화·변주를 선호하며, ③ 온라인을 통한 연대를 추구”[11]한다. 그 밖에도 MZ세대는 온라인을 통해 기념사진을 남기는 것을 즐기고[12], 복잡하고 불편한 과정을 좋아하지 않는다는[10] 특징이 있다.

팬데믹과 비대면 플랫폼 발달에 따라 디지털 트렌드를 주도하는 “MZ세대”는 온라인에서의 만남이 일상화되었다. 수업, 소모임은 물론 “랜선 모임”이라고 불리는 친목 모임이나 회식도 비대면 소통 플랫폼을 통해 진행된다. 통계자료에 따르면 Z세대 청소년의 약 92%는 랜선 모임에 참여해본 경험이 있으며, 73%는 현재 1~2개의 랜선 모임에 참여 중인 것

으로 밝혀졌다 [13]. 이들은 디지털 환경이 익숙하므로 비대면 소통을 대면 소통이라고 생각하는 비율이 높아[3], 비대면 플랫폼을 이용한 소통에 상당히 긍정적인임을 알 수 있다. 이들은 또한 여행에 대한 갈증을 “랜선 여행”이라는 새로운 디지털 콘텐츠 문화를 만들어내고 공유함으로써 충족시켰다. 랜선 여행은 유명 관광지에 온라인으로 방문하고 그 장소에 자신의 모습 혹은 지인의 모습을 합성한 뒤 기념사진의 형태로 SNS에 게시하는 문화이며 유명인들이 동참하여 인기를 끌었다.

## 2-2 영상처리 및 영상이해 기술

본 장에서는 제안하는 영상 합성 서비스를 구성하기 위한 핵심 요소 기술들에 대해 논의한다. 해당 요소 기술들은 인공지능 기술 개발 또는 기계학습을 통해 개발될 수 있다. 먼저 인공지능이란 인간의 사고 및 판단 능력을 컴퓨터 알고리즘으로 구현하여 다양한 곳에 활용하는 기술이다[14]. 인공지능의 용어 및 개념은 1956년에 처음 정의되었으며 2000년대 중반 정보통신기술의 발전으로 생성된 빅데이터와 함께 크게 발전하였다. 최근에는 인공지능을 학습하기 위한 기계학습 기법뿐만 아니라 딥러닝 기술의 비약적인 발전에 따라 새로운 인공지능 붐이 시작되었다[15]. 딥러닝은 심층 인공신경망을 학습시켜 대규모의 데이터나 복잡한 데이터에서도 필요한 기능을 고성능으로 수행하는 기계학습 알고리즘으로서 특히 이미지 분류, 이미지 분할, 객체 탐지 등의 컴퓨터 비전 응용 기술들의 성능을 매우 발전시켰다.

본 연구에서 활용하는 영상 처리 및 영상 이해 기술은 크게 세 가지이다: ① 객체 검출, ② 영상 분할, ③ 영상 생성. 우리는 해당 요소기술들을 기반으로 비대면 소통 플랫폼에 적합한 서비스를 개발한다. 먼저 객체 검출 기술은 주어진 입력 영상에서 원하는 객체의 위치와 크기를 예측하는 기술이다. 최근 딥러닝 기술을 발전으로 높은 성능의 객체 검출 기술들이 많이 등장했다[4,16,17]. 다양한 딥러닝 기술 기반의 객체 검출 기술 중 특히 YOLO[4]는 이미지 내의 경계 상자(Bounding box)를 다차원적으로 분리하고, 클래스 확률 예측을 하나의 회귀 문제로 정의해 객체 검출을 매우 효과적으

로 수행한다. YOLO는 네트워크의 구조 및 구성요소가 복잡한 기존의 Region-based CNN [17] 기법보다 속도가 빠르며, 성능 또한 우수하다. 다음으로 영상 분할 기술은 주어진 입력 영상의 모든 픽셀에 대한 클래스를 예측하는 문제이다. 고전적인 영상 분할 기술은 영상에서 물체의 경계 부분을 검출하거나 관심 물체와 배경을 구분하는 단순한 이진 클래스 예측 문제였다[5]. 최근 딥러닝의 발전으로 수십에서 수백 개의 클래스를 예측하는 의미론적 영상 분할 기술[6]들이 개발되고 있다. 해당 기술은 자율주행 자동차, 의료영상 분석, 실내/외 영상이해 등 다양한 응용 분야에 활용되고 있다.

마지막으로 영상 생성 기술은 가상의 영상을 생성하거나 입력된 영상을 원하는 형태로 변환해주는 응용에 활용된다. 특히 해당 기술은 적대적 생성 신경망 (Generative Adversarial Networks, 이하 GAN) 구조를 응용하여 다양하게 개발되고 있다[7]. GAN은 생성자와 구분자가 서로 적대적으로 경쟁하며 네트워크를 학습한다. 최근 수많은 데이터 및 정교한 모델의 발전으로 인간이 구분하기 어려운 수준의 가상 데이터를 생성하고 있다. 그뿐만 아니라 GAN은 다양한 연구에서 특정 화가의 화풍을 학습해 입력 영상의 스타일을 성공적으로 변환하였다[8]. 앞서 서술한 다양한 영상처리 및 영상이해 기술들은 개별적으로 개발되어있으며 비대면 플랫폼에 그대로 적용이 어렵다. 본 연구에서는 기존 기술들을 개선하고 하나의 프레임워크로 완벽히 통합하여 새로운 형태의 디지털 콘텐츠 제작 서비스를 제안한다.

2-3 관련 연구 및 서비스

본 연구가 제안하는 “비대면 소통 플랫폼 맞춤형 자동 영상 합성 서비스”에 관련된 선행 연구는 다음과 같다. 먼저 관심 객체 분리하고, 원하는 배경 영상을 합성하는 기술[18][19]에 대한 연구가 이뤄지고 있다. 해당 연구들에서는 의미론적 영상분할[6] 기술을 기반으로 영상에서 관심객체만을 분리해내며 새로운 배경 이미지를 단순 삽입한다. 그러나 이러한 기존 기술들은 전경과 합성을 위한 후경 영상간의 밝기, 해상도 등의 정보가 상이할 경우, 합성의 결과물이 부자연스러운 단점이 있다. 이러한 문제점을 해결하기 위해 영상 합성의 자연스러움을 키워 현실성을 향상하기 위한 연구[20]가 진행되었다. 해당 연구는 입력 영상에 대한 색조, 휘도와 같은 표준 통계적 척도를 평가 및 조절함으로써 자연스러운 이미지를 연출한다. 그러나 해당 기술은 입력과 배경 영상들의 수치들을 자동으로 조정할 때 전경과 배경 사이에 큰 밝기 차이가 있으면 빈번한 연산으로 인해 오버헤드가 발생하는 문제점이 있다.

전후경 분리 및 합성, 자연스러운 이미지 생성과 같은 기술들에 관해서는 선행 연구가 존재하였으나, 본 연구에서 목표로 하는 다수의 개별 객체를 자동으로 합성하는 기술에 관한 연구는 이뤄지고 있지 않았다. 보통 카메라 성능에 따라 이미지 상에서의 사람 객체의 크기 및 균일화되지 않은 모습, 출력된 이미지 해상도 등이 다르다.

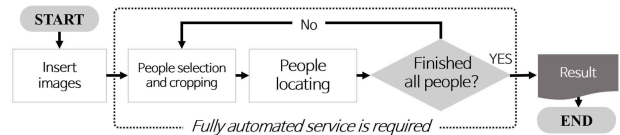


그림 3. 기존 영상 합성 프로그램을 이용한 단체사진 합성 절차  
 Fig. 3. A process of making a group photo through existing image editing programs

또한, 이미지 내에 얼굴과 같은 정해진 특정 구역에 이미지를 합성하는 것이 아닌, 여러 의미론적 정보가 담겨있는 배경에 이미지를 합성하는 것이기 때문에 어려움이 있다. 따라서 각각 분리된 형태의 사람들이 존재하는 이미지에 관해 자동으로 한 장의 이미지로 합성하는 기술에는 다양한 변수가 존재하기에 일반적인 자연스러운 사진처럼 출력하는 기술적 구현이 어렵다.

본 연구에서 중점적으로 다루고 있는 일반적인 비대면 소통 플랫폼은 이용자들의 영상이 동일한 규격의 칸 안에 한 명씩 정렬된 형태로 구성된다(그림 1.(좌)). 이는 “랜선 모임” 등을 진행하기에는 적절한 플랫폼이지만 기념사진으로 남기 기에는 자연스러움과 재미가 떨어진다. 이는 재미를 추구하고, 트렌드를 소비하는 MZ세대의 요구를 충족시키지 못한다. ZOOM 애플리케이션에서는 유료 라이선스를 구매하면 회의실 형태로 프로그램 화면을 구성해주지만, 사용자 영상이 각각 위치한 기본 규격에서 크게 벗어나지 않는다. 그로 인해 사용자는 비대면 소통 플랫폼에서 캡처한 기념사진을 자연스러운 영상으로 합성하기 위해 기존의 상용 영상 합성 프로그램 이용을 이용해야 한다. 기존의 상용 영상 합성 프로그램은 입력한 사람 수에 비해한 수많은 반복 작업과 다양한 단계가 필요하다(그림 3). 그 뿐만 아니라 프로그램을 활용하는 사용자의 전문성에 따라 그 결과물의 질이 일정하지 않을 수 있다. 이렇듯 기존의 서비스들은 빠르고 간편한 것을 추구하는 MZ세대의 요구를 충족시키지 못한다.

III. 제안 기술 및 서비스

3-1 제안 서비스의 전체 구조 및 목표

본 연구에서 제안하는 영상 합성 기술 및 이를 바탕으로 한 서비스의 전체 구조는 그림 2.와 같다. 사용자는 웹 페이지를 통해 간편하게 본 서비스를 이용할 수 있으며 서비스의 입력으로는 “다인이 포함된 한 장의 캡처 사진”만이 있다. 웹 페이지에서 사용자 입력을 받으면 해당 정보는 메인 서버로 전송되며 입력 사진의 크기에 상관없이 동작하도록 구현하였다. 모든 알고리즘의 동작은 메인 서버에서 처리되며 메인 서버 구현은 구조가 단순하고 딥러닝 알고리즘의 적용이 용이한 과이션 웹 프레임워크 Flask를 활용하였다. 또한, 이어지는 연구에서 제안하는 기법들을 통해 메인 서버에서 다양한 스

타일의 단 사진을 합성하고 이를 웹 페이지에 출력한다. 사용자는 최종적으로 본인이 원하는 양식의 이미지를 선택 및 저장할 수 있으며 다양한 타 매체 (SNS)에 공유가 가능하다.

제안 서비스에서 목표로 하는 것은 크게 두 가지이다. 첫째 목표는 캡처 영상에 등장하는 사람 수와 관계없이 자연스러운 합성 영상을 생성하는 것이다. 이를 위해서는 사람의 상반신 및 얼굴의 위치를 정확히 탐지해야 하며 여러 인물을 올바른 곳에 위치시키는 알고리즘 개발이 필수적이다. 두 번째 목표는 최소화된 사용자 입력이다. 이를 위해 제안 서비스에서는 사용자 입력을 제외한 모든 부분을 자동화하였다.

### 3-2 사람 검출 및 배경 분할

자동화된 서비스를 구현하기 위해서는 입력 영상에 등장한 사람들을 컴퓨터 알고리즘이 자동으로 검출해야 한다. 이를 위해 본 연구에서는 구조가 가벼워 속도가 빠르고 더불어 높은 정확도를 가지는 YOLO version 5 (이하 YOLO) [21] 객체 검출기를 기반 기술로 활용한다. 해당 객체 검출기에서 사람 클래스와 얼굴 클래스에 해당하는 검출 결과를 활용한다.

얼굴 클래스(Face class)는 어떠한 학습 영상에서도 일관적인 외형 정보를 가지기 때문에 기존 공개 모델로 충분히 검출된다. 반면 활용하는 객체 검출 모델은 사람의 전신 영상이 대부분인 MS coco [22] 및 ImageNet [23] 데이터 셋을 기반으로 학습되어있다. 즉, 상반신만 노출되는 비대면 소통 플랫폼에서는 기존 모델의 사람 클래스에 대한 검출 성능이 떨어지는 한계가 있다. 이를 단순 개선하기 위한 한 가지 방법으로, 사람 클래스로 판단하는 모델의 신뢰 임계값을 낮출 수 있다. 그러나 이 경우 사람 검출은 성공하지만, 사람이 아닌 영역을 사람으로 잘못 판단하는 거짓 양성 (False Positive) 응답이 다수 발생할 위험이 있다. 본 연구에서 목표로 하는 자연스러운 영상 합성을 위해서는 사람 검출에 실패하거나 거짓 양성이 발생해서는 안 된다.

이를 극복하기 위해 우리는 ① 상반신만 포함된 사람 영상 데이터를 대규모로 수집하고 추가로 ② 비대면 플랫폼 환경의 배경 영상 (사무실, 카페, 방 등) 데이터를 다양하게 수집하였다. ①에 해당하는 데이터는 앞서 설명한 비대면 플랫폼 영상에서 사람 검출 실패 위험을 줄이기 위해 수집하였으며 ②에 해당하는 데이터는 거짓 양성이 발생할 위험을 줄이기 위해 사람이 없는 배경 영상 위주로 수집하였다. 수집된 데이터 ①, ②는 모두 기존 모델 강화를 위한 미세 조정 (Fine-tuning)에 활용된다. 모델의 미세 조정은 기존 모델의 가중치 값을 초기값으로 설정한 뒤 추가 데이터에 대해 반복적으로 재학습을 수행하는 방식으로 진행한다. 이러한 모델 미세 조정은 단순하지만 새로운 도메인의 데이터에 적응시키는 데 매우 효과적이다.

객체 검출기의 검출 결과는 6차원의 벡터로 다음과 같이 표현되며 그림 4.(좌)와 같이 영상에 표시된다.

$$[c_i, x_i, y_i, w_i, h_i, s_i], \text{ where } i = 1, 2, \dots, N, \quad (1)$$

$c$ 는 검출 객체의 클래스,  $x$ 와  $y$ 는 객체의 중심 픽셀 좌표 값,  $w$ 와  $h$ 는 객체의 가로와 세로 길이를 각각 나타내며,  $s$ 는 검출 신뢰 점수(0~1)를 의미한다.  $i$ 는 검출 결과의 인덱스를 나타내며  $N$ 은 검출된 전체 인물의 수를 나타낸다. 본 연구에서는 사람 클래스 중 검출 신뢰 점수가 충분히 높은 ( $s \geq 0.8$ ) 결과들에 대해서만 최종 검출 결과로 인정하며 그 밖의 응답들은 모두 사전에 제거된다. 또한, 사람 상반신 검출 결과와 얼굴 검출 결과는 인덱스  $i$ 를 공유하여 일대일로 대응된다. 따라서 입력 영상에 대한 상반신 길이  $u_i$ 와 그에 대응되는 인물의 얼굴 길이  $f_i$ 는 다음과 같이 나타낼 수 있다.

$$u_i = [w_i^u, h_i^u], \quad (2)$$

$$f_i = [w_i^f, h_i^f], \quad (3)$$

$u_i$ 는 수식 (1)의  $c=0$  (상반신)에 해당하는 길잇값들,  $f_i$ 은 수식 (1)의  $c=1$  (얼굴)에 해당하는 길잇값들이다.

검출된 사람 영상들은 각자의 공간에서 다양한 배경으로 촬영되었으므로 자연스러운 영상 합성을 위해 사전에 배경을 제거해야 한다. 이를 위해 PointRender 신경망 기반으로 빠르고 정확한 이미지 분할이 가능한 PixelLib[24]를 활용하였다. 본 연구에서는 사람만을 분할하기 때문에 사람은 전경, 나머지는 배경으로 간주하고 모두 제거한다. 최종적으로 그림 4.(우)와 같은 배경이 제거된 인물 영상들을 각각 얻게 되며 해당 영상들은 3-3장에서 제안하는 기법들을 통해 합성된다.

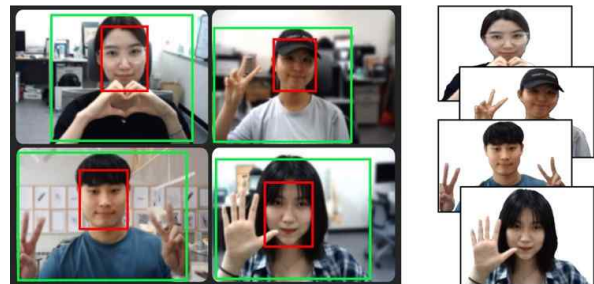


그림 4. (좌) 사람 상반신 (초록 상자) 및 얼굴 (빨간 상자) 검출 결과 (우) 인물 별 배경 제거 결과

Fig. 4. (Left) Upperbody (green box) and face (red box) detection results (Right) Background removal results of each person

### 3-3 자연스러운 합성 영상 생성 기법

비대면 소통 플랫폼에 등장하는 사람들의 영상을 촬영한 카메라의 환경 (내/외부 카메라 인자)이 서로 다르므로 영상들의 스케일이 일정하지 않다. 이 경우 별다른 규칙 없이 임의로 사람 영상들을 위치시키면 인물들끼리 서로 가리거나

크기가 제각각 이어서 매우 부자연스럽다. 본 연구에서는 위와 같은 어려움을 극복하고 자연스러운 단체 영상을 합성하기 위해 다음과 같은 제약 조건들을 둔다. ① 각 인물 간의 크기 차이가 크지 않도록 한다. ② 인물들 간 얼굴이 겹치지 않으며 상반신이 화면 밖으로 나가지 않도록 한다.

본 연구에서는 먼저 사용자 영상들의 크기를 서로 유사하게 정규화한다. 이를 위해서는 다음과 같이 전체 등장인물의 가로 얼굴 길이의 평균을 먼저 계산한다.

$$F = \frac{1}{N} \sum_{i=1}^N w_i^f, \quad (4)$$

구해진 평균 얼굴 크기를 이용하면 각 인물의 얼굴과 평균 길이 간의 비율 인자  $r_i = F/w_i^f$  을 각각 계산할 수 있으며 해당 값을 기존 얼굴 길이에 곱해주면 검출한 모든 얼굴의 가로 길이를  $F$ 로 정규화할 수 있다. 그러나 합성할 인물들의 얼굴 길이가 모두 동일해지면 오히려 부자연스러울 수 있다. 이러한 문제를 극복하기 위해 평균값 1, 분산 값이 0.01인 표준 정규 분포( $\alpha_i \sim N(1,0.01)$ )에서 임의의 값  $\alpha_i$ 들을 샘플링 한다. 샘플링 되는  $\alpha_i$  값들은 분포의 3-sigma 범위에 대부분 해당하는 0.8에서 1.2 사이 값들이 주로 추출된다.

최종적으로 비율 인자  $r_i$ 와 임의 추출 값  $\alpha_i$ 를 곱한  $r_i \alpha_i$ 을 ‘사람 간 정규화 인자’로 정의한다. 해당 정규화 인자를 수식 (2)와 (3)에 곱해 정규화된 인물 상반신 및 얼굴 길이를 다음과 같이 새롭게 얻는다. 해당 정규화를 통하면 각 사람들간의 크기 차이를 좀 더 일관적으로 만들 수 있는 효과가 있다.

$$\bar{u}_i = r_i \alpha_i [w_i^u, h_i^u] = [\bar{w}_i^u, \bar{h}_i^u], \quad (5)$$

$$\bar{f}_i = r_i \alpha_i [w_i^f, h_i^f] = [\bar{w}_i^f, \bar{h}_i^f]. \quad (6)$$

길이가 정규화된 사용자 영상들은 최종적으로 입력한 배경에 배치해야 한다. 이때 배경 영상은 가로 및 세로 크기 각각  $W, H$ 인 임의의 크기로 주어지기 때문에 사용자 영상 크기와 배경 크기 간의 비율 조정이 한 번 더 필요하다. 본 연구에서는 인물들이 영상 밖으로 나가지 않으면서 최대한 크게 합성하고자 한다. 이를 위해서 정규화된 인물들 상반신의 가로 길이 합을  $W^u = \sum_{i=1}^N \bar{w}_i^u$ , 와 같이 계산하고 배경 영상과의 비율인  $q = W/W^u$ 를 ‘합성 정규화 인자’로 정의한다.

제안한 방법으로 합성한 단체 사진은 위치 및 구도는 자연스럽지만 각 사람 영상이 서로 다른 공간 및 환경에서 촬영되었기 때문에 조명 및 색감은 여전히 제각각이고 부자연스럽다. 최종적으로 모든 인물 및 합성 배경이 서로 어우러지는 단체 합성 사진을 생성하기 위해 합성된 영상의 스타일을 변화한다. 이를 위해 다양한 스타일 변환 기술 [7,8] 등이 활용될 수 있으며 본 연구에서 제안하는 서비스에는 카툰스타일

[25] 변환 기술을 적용했다. 최종적으로 합성된 단체 사진은 그림 9.와 같이 합성 인물의 수 및 크기에 상관없이 상당히 자연스러운 것을 확인할 수 있다.

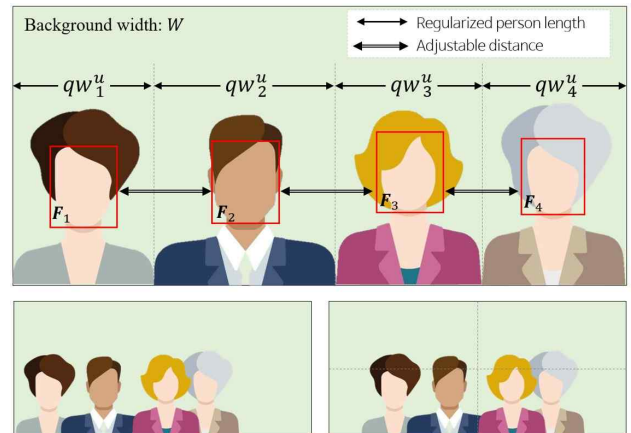


그림 5. (상) 인물 크기 정규화 및 배치 (좌 하단) 인물 간 재배치 (우 하단) 최종 배치 결과물 예시  
**Fig. 5.** (Top) Size normalization and location arrangement (Left Bottom) Relocation between people (Right Bottom) Final alignment result

해당 정규화 인자를 수식 (5) 및 (6)에 곱하면 합성을 위한 인물들의 최종 크기가 결정되며 그림 5.(상)와 같이 인물들을 가로로 나열했을 때 서로 겹치지 않으면서 화면에 가득 차는 크기로 정규화가 된다. 합성된 인물들의 상반신 및 얼굴의 경계 상자 (위치 및 크기)  $U_i$ 와  $F_i$ 는 다음과 같이 정의된다.

$$U_i = [x_i^u, y_i^u, \bar{w}_i^u, \bar{h}_i^u], \quad (7)$$

$$F_i = [x_i^f, y_i^f, \bar{w}_i^f, \bar{h}_i^f]. \quad (8)$$

그러나 모든 인물을 일정한 간격으로 배치하면 오히려 부자연스럽다. 자연스러운 연출을 위해 각 인물의 가로 중심 좌표  $x_i^u$ 에 약간의 조정을 한다. 이때 인물별 중심좌표의 조작에는 사람들의 얼굴 경계상자  $F_i$ 가 서로 겹치지 않는 범위로 한정한다 (그림 5.(좌 하단)). 위치 조정이 완료되면 최종적으로 모든 인물의 가로 중심 좌표의 평균값을 합성 영상의 중앙에 위치하도록 보정 하여 인물들이 영상 한쪽으로 치우치지 않게 한다. 그 결과 그림 5.(우 하단)와 같이 인물들이 자연스럽게 배치된 단체 합성 사진을 얻을 수 있다.

### 3-4 사용자 인터페이스

제안하는 서비스에서는 사용자의 접근성이 좋고 간편하게 이용할 수 있는 웹을 사용하였다. 웹 페이지의 뼈대는 HTML, 디자인적 요소는 CSS를 이용하였다. 웹 서비스 프레임워크(Framework)는 Flask를 사용하였다.

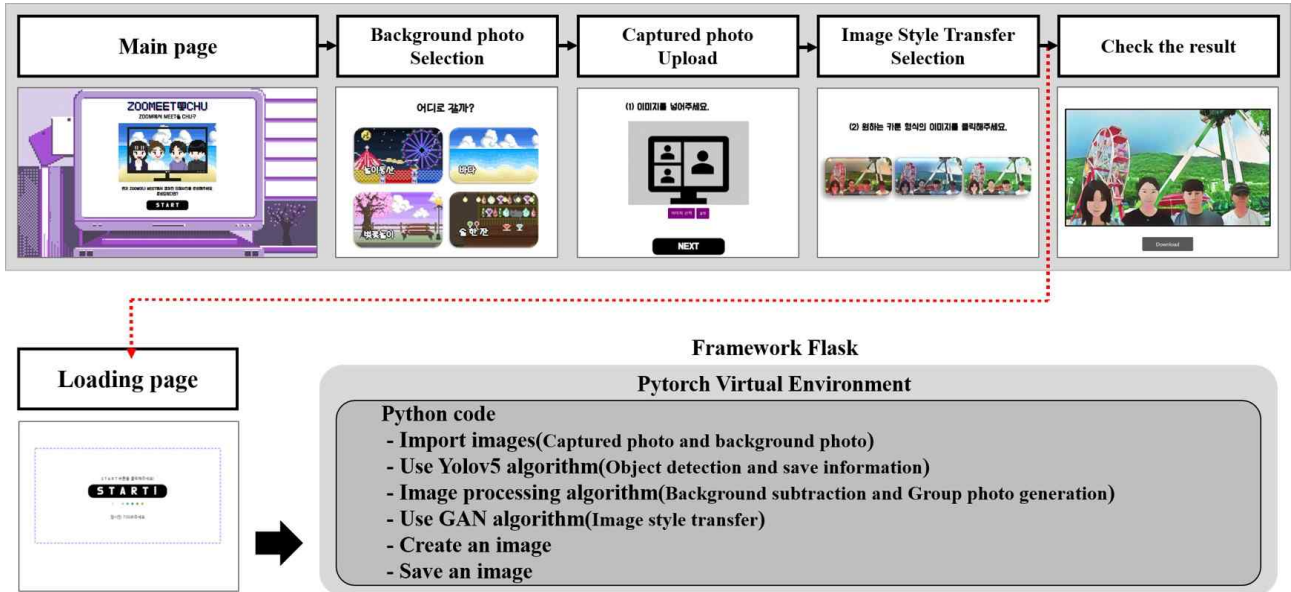


그림 6. (상) 제안 서비스의 동작 프로세스 및 사용자 인터페이스; (하) 이미지 처리 환경  
 Fig. 6. (Top) Process of proposal service and user interface; (Bottom) Image processing environment

웹 서비스에 존재하는 파이썬(Python) 기반의 다양한 프레임워크 중 Flask는 매우 가볍고 단순한 구조를 가진다. 그뿐만 아니라 활용 자유도가 높아, 사용자가 원하는 라이브러리와 패키지를 연동할 수 있는 장점이 있다.

사용자 인터페이스는 그림 6.(상)과 같다. 사용자가 웹 서비스에 접속한 후 시작(Start) 버튼을 클릭하면 다음 페이지로 넘어간다. 이어지는 페이지에서 사용자는 기본적으로 제공되는 4개의 배경 사진 중에 하나를 선택할 수 있다. 선택 후 비대면 소통 플랫폼에서 캡처한 사진을 웹에 업로드한다. 이후 카툰 형식 선택 페이지에서 변환하고자 하는 이미지 스타일을 선택한다. 이때, 사용자는 3개의 이미지 스타일 중 결과 이미지를 눈으로 확인하고, 선호하는 스타일을 선택할 수 있다. 선택 후 바로 로딩 페이지로 넘어가는데, 이때 사용자가 시작 버튼을 클릭하면 선택한 배경 이미지, 비대면 소통 플랫폼에서 캡처한 사진, 이미지 변환 스타일을 토대로 결과 이미지를 생성한다.

시작 버튼 클릭 이후 작동되는 과정은 그림 6.(하)와 같다. 로딩 페이지에는 플라스크 프레임워크 사용하여 Pytorch(파이토치) 기반의 가상환경이 생성되어있다. Pytorch는 풍부한 손실 함수를 사용할 수 있으며, 효율적인 계산으로 가장 많이 사용하고 있는 딥러닝 프레임워크이다. 그 안에는 Python을 이용해 다음과 같이 구현되어있다. 먼저 사용자가 업로드하여 시스템 내에 저장된 이미지를 가져온 뒤, YOLO를 사용해 사람 객체를 찾아 관련 정보를 저장한다. 저장된 정보를 이용해 배경 삭제 및 합성과 같은 이미지 처리 과정을 거친다. 이후 GAN을 사용해 이미지의 스타일을 변환시킨다. 최종적으로 한 장의 이미지가 생성되고, 생성된 이미지를 저장한다. 다음 페이지로 이동된 웹에서 사용자는 본인이 원하는 합성 결과를 확인할 수 있으며, 해당 이미지를 다운 받거나 SNS에 공유할 수 있다.

서비스의 디자인을 위해서는 MZ세대의 특성을 고려하였다. 그들과 같은 젊은 소비자들에서 과거의 콘텐츠들을 재해석하여 새로운 이미지를 창출하는 ‘뉴트로’ 현상이 인기를 얻고 있다 [26]. 특히 그중에서도 픽셀아트를 이용한 홈페이지나, 제품들이 많은 호평을 얻었다. 이러한 부분을 참고하여 MZ세대의 요구에 맞춘 픽셀아트들은 모두 직접 디자인하였으며 모든 서비스에 해당 디자인들을 적용하였다.

#### IV. 실험 결과

##### 4-1 MZ세대를 위한 영상 합성 서비스 설문

실험에 앞서 본 연구에서는 MZ세대를 위한 영상 합성 서비스에 관한 설문조사를 진행했다. 설문은 2021년 10월 5일부터 10월 10일까지 6일간, 총 134명의 MZ세대에 해당하는 사람들을 대상으로 하였다. 먼저 본 설문에서는 ‘MZ세대의 특징’에 대해 설문하였다. 표 1. 에 따르면 ‘재미를 추구하는 MZ세대’의 비율은 전체 응답자의 86.6%, ‘복잡하고 불편한 과정을 선호하지 않는 MZ세대’의 비율은 전체 응답자의 65.7%이며 ‘기념사진 촬영을 즐기는 MZ세대’의 비율은 전체 응답자의 56.7%인 것으로 각각 확인되었다. 따라서 본 연구에서 제안한 서비스의 목표가 MZ세대의 실제 요구와 부합함을 알 수 있다.

또한, 진행된 설문에서는 MZ세대의 랜선 모임 참여 경험 비율과 경험한 랜선 모임 활동에 대해 설문하였다. 그림 7.와 같이 약 84%가 랜선 모임에 참여했으며 랜선 모임에 현재 참여 중인 응답자 비율은 약 70%인 것으로 확인되었다. 랜선 모임에 참여한 경험이 있는 사람들 중 비대면 소통 플랫폼을

통해 소모임을 가진 응답자의 비율은 69.9%이며, 술자리 비율은 24.8%인 것으로 확인되었다 (표 2). 비대면 소통 플랫폼을 MZ세대들은 상당히 자주 접하며 친숙함을 알 수 있다.

마지막으로 비대면 소통 플랫폼 맞춤형 영상 합성 서비스 (이하 ZOOMEEATCHU)의 이용 의향을 평가했다. 그림 7. 에서 확인할 수 있듯, ZOOMEEATCHU를 이용하겠다는 비율은 전체 응답자 중 약 68%로 확인되었다. 응답자의 약 3명중 2명이 해당 서비스에 대해 긍정적으로 생각하고 있으며, 본 연구의 서비스는 디지털 콘텐츠 문화를 주도하는 MZ세대의 기대를 상당히 충족하는 디지털 콘텐츠임을 알 수 있다. 단 약 32%의 설문 참여자에 대해 서비스 이용에 대한 긍정적인 의사가 없었으며 이를 극복하기 위한 서비스 보완 및 방향성 설정은 향후 과제로 남아있다.

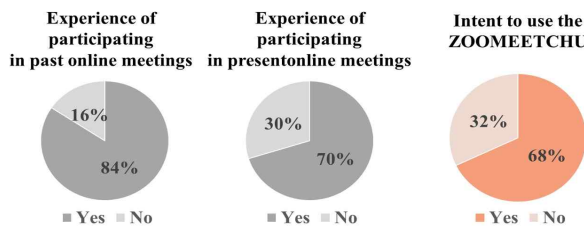


그림 7. (설문결과) 비대면 소통 플랫폼 이용 및 모임 경험; 제안 서비스의 사용 의사

Fig. 7. (Survey result) Experiences of using online meeting platform and preference for ZOOMEEATCHU

표 1. (설문결과) MZ세대 특징

Table. 1. (Survey result) Characteristics of Generation MZ  
A total of 134 people (Duplicate Check)

Characteristics of Generation MZ	Respondents
Pursuing for Fun	116 86.6%
Hates complicated and uncomfortable things	88 65.7%
Prefer to take Pictures	76 56.7%
Thinking freely	74 55.2%
Speed	62 46.3%
Pursuing for Solidarity Online	47 35.1%
N/A	1 0.7%

표 2. (설문결과) MZ세대가 경험했던 랜선 모임 활동

Table. 2. (Survey result) Online meeting activities experienced by the generation MZ  
A total of 113 people (Duplicate Check)

Online meeting activities	Respondents
Lecture	110 97.3%
Small club	79 69.9%
Work	46 40.7%
Club on campus	42 37.2%
Drinking party	28 24.8%
Study	5 4.4%
Other(Blind date, Game, ...)	5 4.4%

#### 4-2 사람 검출기 개선

YOLO v5 (이하 YOLO)을 학습하기 위해 활용 가능한 데이터셋 (MS coco [22], ImageNet [23])에는 사람의 전신 영상 데이터가 대부분이다. 이를 통해 사람 검출기를 학습하면 상반신만 노출되는 비대면 소통 플랫폼에서는 검출 성능이 떨어져 일부 사람을 검출하지 못하는 상황이 일어난다. 사진 한 장에서 일부만 검출하게 된다면 검출되지 않은 사람들은 사진에 합성시킬 수 없을 것이다. 이를 개선하기 위해 그림 8.(상)과 같이 사람의 상반신만 포함된 사람 데이터 700장, 그림 8.(하)와 같이 사람이 없는 비대면 플랫폼 환경의 배경 (사무실, 카페, 방 등) 데이터 170장을 추가로 수집하였다. 그 뿐만 아니라 기존 공개 데이터 셋 MS coco[22] 중 사람의 앞모습 영상 위주로 2,200장을 활용하였다. 이는 비대면 소통 플랫폼에 참여한 사람은 정면 - 상반신이 노출됨을 반영하기 위함이다.

수집된 데이터는 Train : Valid = 8 : 2 비율로 기존 모델 강화를 위한 미세 조정(Fine-tuning)에 활용하였으며 개선된 사람 검출 모델을 만들었다. 다양한 비대면 플랫폼 환경의 단체 영상을 100장 수집했으며 해당 영상에는 총 510명이 등장한다. 이에 대한 사람 검출 실험 결과, 표 3.과 같이 기존 모델의 정밀도 성능을 유지하면서 재현율 (Recall)을 5.3%를 향상하였다. 객체 검출을 위한 신뢰도 임계값을 본 연구에서는 0.8로 설정 하였으며 해당 파라미터는 필요에 따라 조정가능 하다. 제안하는 프레임워크를 위한 얼굴 검출은 기존의 YOLO 모델의 성능이 충분하여 그대로 활용하였다.

#### 4-3 단체 영상 합성

제한한 서비스의 입력 영상으로 이용된 영상은 비대면 소통 플랫폼의 일반적인 규격을 따른다. 다시 말해 인물들이 적절히 촬영되었다면 어떤 종류의 플랫폼에서 촬영되었더라도 본 서비스에서 활용할 수 있다. 비대면 소통 플랫폼에서 캡처한 영상 한 장을 서비스에 업로드하면 자연스러운 합성 영상이 생성되며 입력된 사람의 인원수와 상관없이 각 인물의 위치가 자동으로 추천된다. 그림 9.는 최대 다섯 명에 대한 합성 사진 결과물을 보여주며 인물들이 적절히 배치된 것을 확인할 수 있다. 실제 시스템 구현을 위해서는 배경별로 추천하는 합성 인물 최대 수를 설정하였으며 입력 영상으로 너무 많은 수의 사람이 들어오는 경우 서비스에서는 무리하게 합성하지 않고 사용자에게 적절한 피드백을 주도하도록 하였다. 또한, 제한한 서비스에서는 각각의 장소에서 촬영된 인물들 및 배경이 하나로 합성했을 때 자연스러운 연출을 위해 이미지 스타일 변환 기술을 적용하였다.

그림 9는 최종 결과물로 하나의 예시만 보여주고 있지만, 실제 서비스를 이용하는 사용자는 다양한 결과물을 선택할 수 있다. 본 서비스에서 적용한 스타일 변환 기술뿐만 아니라 다른 종류의 기술도 사용자 기호에 맞춰 적용할 수 있다.



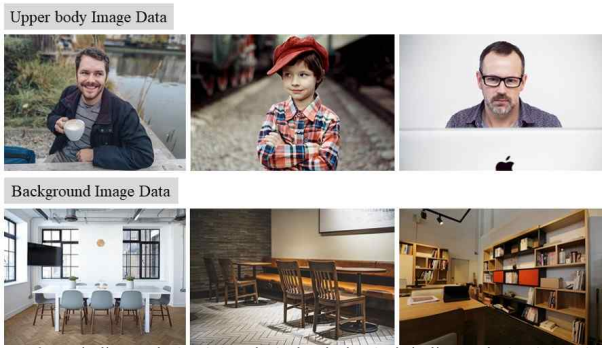


그림 8. (상) 수집된 학습 상반신 영상 예시 (하) 수집된 학습 배경 영상 예시

Fig. 8. (Top) Examples of additional upperbody Images (Bottom) Examples of additional background Images

표 3. 개선 전후 YOLO v5 모델의 성능 비교  
Table 3. Comparison to Yolo v5 models

Person Detection Model	Confidence Threshold of 0.8	
	Recall	Precision
Yolo v5 (baseline)	86.1 %	100 %
Yolo v5 (improved)	91.4 %	100 %

## V. 결론

본 연구에서는 비대면 소통 플랫폼에서 MZ세대의 요구를 충족시키는 자연스러운 단체사진 연출이 어렵다는 문제점을 제시하였으며 이를 고려한 디지털 콘텐츠 제공 서비스를 새롭게 제안하였다.

이를 위해 기존의 최신 인공지능 연구 기술들을 적절하게 활용하였으며 자연스러운 단체사진 연출이 가능하도록 새로운 알고리즘을 제안하였다. 제안한 서비스는 비대면 플랫폼의 종류에 종속적이지 않으며 웹을 통한 간편한 활용이 가능하므로 높은 이식성을 보인다. 그뿐만 아니라 다양한 실험 영상에 대해 제안 서비스가 올바르게 동작함을 확인하였다. 제안 서비스는 본 연구에서 예시로 든 소규모 모임 외에도 온라인 입학식, 학회 및 세미나와 같은 대규모 행사에서도 자유로운 활용이 가능하다. 이를 통해 팬데믹 상황에서 비대면 모임을 장려하고 많은 사람에게 활력과 재미를 줄 수 있기를 기대한다.

본 연구에서는 3-3장에서 기술한 것처럼 사용자의 조작을 최소화하여 자동으로 단체 사진을 합성하는 것을 목표로 하였다. 그러나 서비스에서 사용자에게 더 다양한 합성 결과물들을 추천하고 최종 결정을 사용자가 선택하도록 변경할 수도 있을 것이다. 물론 이 경우 사용자 조작이 늘어나는 단점이 있지만 합성 사진 배치에 보다 자유도를 높일 수 있는 장점이 있다. 또한 추후 연구로써 사용자가 선택한 배경의 장면 정보를 이해하고 [6] 이해된 정보를 활용하여 더욱 자연스러운 인물 배치를 할 수 있을 것이다. 예를 들면 배경에 등장하는 사물들을 인지해 해당 사물과 상호작용하는 형태로 인물들을 배치하거나 장면 전체의 분위기를 파악해 그것과 어울리는 형태로 배치할 수 있을 것이다.

## VI. 고찰 및 추후 연구

본 연구는 영상 합성 시 상이한 이미지 정보에 따라 부자연스러운 이미지에 관해 현실성을 향상시키기 위해 인공지능 영상 스타일 변환 알고리즘을 이용하였다.

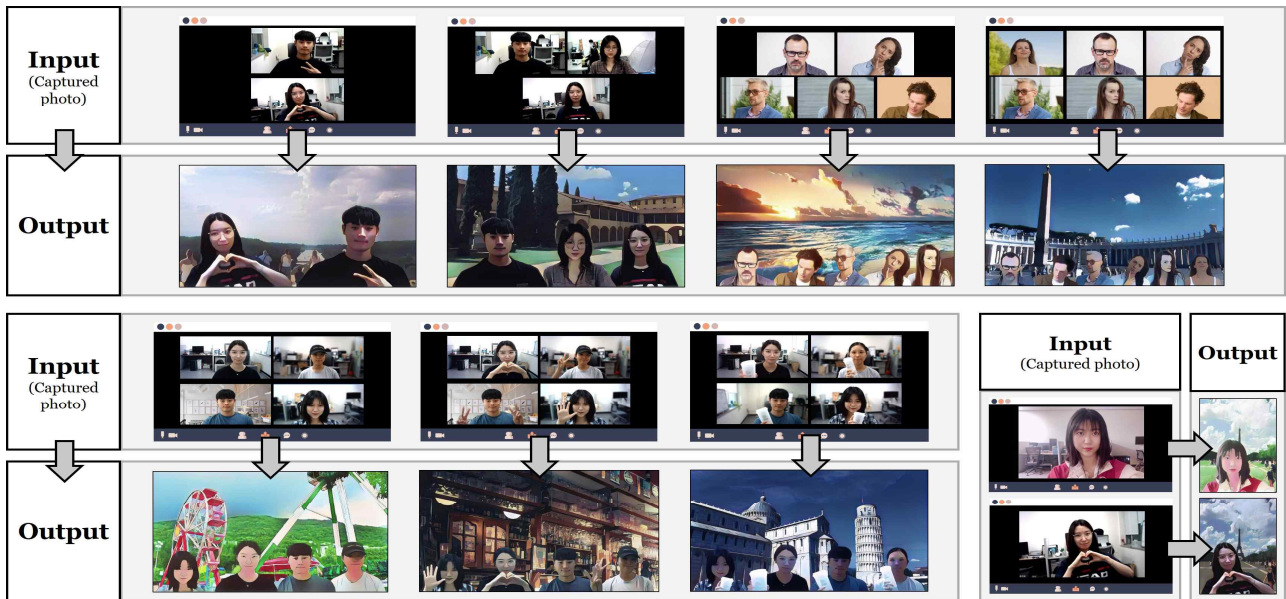


그림 9. 단체 인물 영상 합성 결과 예시

Fig. 9. Examples of group photo synthesizing

그로 인해 기존 유의미한 밝기차이로 인해 빈번한 연산이 발생해 시간적인 오버헤드가 발생했던 통계적 척도를 이용한 자연스러운 이미지 생성의 문제점을 해결하였다. 따라서 배경 이미지 정보와 합성되는 다양한 상반신 사람 객체의 밝기, 해상도와 같은 이미지 정보가 달라도, 자연스러운 이미지 생성이 가능하다. 또한, 선행 연구가 존재하지 않았던 다수의 객체를 자동으로 합성하는 기술에 관해서는 배경 이미지에 상반신 사람 객체의 얼굴 크기 정보를 이용해 이미지를 정규화함으로써 이미지 크기의 차이를 현저하게 줄여, 더욱 자연스러운 합성 결과물이 발생하도록 유도했다.

또한, 기존의 사용 영상 합성 프로그램이 가지고 있는 단점인 ‘복잡한 반복 작업’에 관해서는 사용자가 비대면 소통 플랫폼에서 캡처한 단 한 장의 사진을 서비스에 업로드하면 배경 이미지에 맞춰 파이션 기반의 프레임워크인 플라스크에서 구현된 코드를 이용해 복잡한 과정을 한 번에 처리함으로써 사용자는 손쉽게 결과물을 받아볼 수 있게 되었다. 이는 복잡한 과정을 추구하는 MZ세대의 요구를 충족시켰다고 볼 수 있다.

## 참고문헌

[1] Dong-a science. Finally, a pandemic declaration... What is a pandemic? [internet]. Available: <https://www.dongascience.com/news.php?idx=35044>.

[2] Central Disaster Management Headquarters. Overview of Social Distancing System. [internet]. Available: <http://ncov.mohw.go.kr/socdisBoardView.do?brdId=6&brdGubun=1>.

[3] Ewha Voice. A party in the COVID-19 pandemic? Generation Z's untact meeting. [internet]. Available: <https://inews.ewha.ac.kr/news/articleView.html?idxno=32451>.

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, Real-Time Object Detection", *In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, Las Vegas: NV, pp. 779-788, June 2016. <https://doi.org/10.1109/CVPR.2016.91>.

[5] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness color, and texture cues", *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 26, No. 5, pp. 530-549, May 2004. <https://doi.org/10.1109/TPAMI.2004.1273918>.

[6] Y. Zhu, K. Sapra, F. A. Reda, K. J. Shih, S. Newsam, A. Tao, and B. Catanzaro, "Improving Semantic Segmentation via Video Propagation and Label Relaxation", *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, California: CA, pp.

8856-8856, June 2019.

<https://doi.org/10.1109/CVPR.2019.00906>.

[7] I. Goodfellow, et al., "Generative adversarial nets", *In Proceedings of the 27th International Conference on Neural Information Processing Systems*, USA, pp. 2672-2680, Dec 2014.

[8] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", *In Proceedings of the IEEE international conference on computer vision*, Venice, pp. 2223-2232, Oct 2017. <https://doi.org/10.1109/ICCV.2017.244>.

[9] Dong-a science. If global warming continues, 90% of the world will be at risk of a 'tropical epidemic' by 2080. [internet]. Available: <https://m.segye.com/view/20210712513919>.

[10] M. Kim and S. H. Hong, "A Study on Communication Methods According to the Characteristics of MZ Generation", *The Treatise on The Plastic Media*, Vol. 24, No. 1, pp. 113-120, Feb 2021. <https://doi.org/10.35280/kotpm.2021.24.1.13>.

[11] H. W. Kim and S. J. Lee, "The future of the financial industry led by generation MZ", *WFRI*, pp. 1-12, 2021.

[12] The Korea Times. The culture of photography in 2030 photos. [internet]. Available: <https://www.hankookilbo.com/News/Read/201502041794920278>.

[13] Busan best economy. 84 percent of teens "raenseon meeting activities." [internet]. Available: <https://www.busanecconomy.com/news/articleView.html?idxno=250937>.

[14] K. Kim, J. Lee, and C. Lee, "Exploratory Analysis of Knowledge Structure and Evolutionary Trajectory in Korean Artificial Intelligence for Effective Technology Policy", *Innovation studies*, Vol. 16, No. 3, pp. 139-172, Aug 2021. <https://doi.org/10.46251/INNOS.2021.8.16.3.139>.

[15] Y. C. Choi, Computer Science in the ICT Convergence Age, *Saengneung Publishing*, pp. 472, 2017.

[16] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection", *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Seattle, pp. 10781-10790, June 2020. <https://doi.org/10.1109/CVPR42600.2020.01079>.

[17] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 38, No.

1, pp. 142-158, Jan 2016.

<https://doi.org/10.1109/TPAMI.2015.2437384>.

- [18] D. H. Kim, S. H. Park and Y. G. Seo, "Automatic Arm Region Segmentation and Background Image Composition", *The Journal of Digital Contents Society*, Vol. 18, No. 8, pp. 1509-1516, Dec 2017. <https://doi.org/10.9728/dcs.2017.18.8.1509>.
- [19] J. C. Yoon, "Real-time Video Matting for Mobile Device", *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol. 11, No. 5, pp. 487-492, Oct 2018. <https://doi.org/10.17661/jkiiect.2018.11.5.487>.
- [20] D. S. Lee, O. K. Ha and Y. K. Jun, "A Survey on Improving Realism for Image Composition", *Proceedings of the Korean Society of Computer Information Conference*, Vol. 25, No. 1, pp. 37-38, Jan 2017.
- [21] Github. Gleen jocher yolov5 [internet]. Available: <https://github.com/ultralytics/yolov5>.
- [22] T. Y. Lin, et al., "Microsoft coco: Common objects in context.", *European conference on computer vision*, Zurich Switzerland, pp. 740-755, Sep 2014. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48).
- [23] J. Dong, W. Dong, R. Socher, L. J. Li, K. Li and F. F. Li, "Imagenet: A large-scale hierarchical image database.", *Conference on Computer Vision and Pattern Recognition (CVPR)*, Florida: FL, pp. 248-255, June 2009. <https://doi.org/10.1109/CVPR14067.2009>.
- [24] Github. PixelLib [internet]. Available : <https://github.com/ayoolafenwa/PixelLib>.
- [25] Y. Chen, Y. K. Lai and Y. J. Liu, "Cartoongan: Generative adversarial networks for photo cartoonization.", *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Utah: UT, pp. 9465-9474, June 2018. <https://doi.org/10.1109/CVPR.2018.00986>.
- [26] J. H. Lee and H. K. Kim, "The Influence of Z Generation and Millennial Generation Consumer's Curiosity and The Need to Belong on Retro Product Preference: Focusing on Mediating Effect of Self-Expansion", *Journal of Consumer Studies*, Vol. 32, No. 3, pp. 95-121, June 2021. <http://doi.org/10.35736/jcs.32.3.5>.

### 황서빈(Seo-Bin Hwang)



2020년~현재 : 전남대학교 AI융합대학  
IoT인공지능융합전공  
학사과정

※ 관심분야 : 컴퓨터 비전(Computer Vision), 인공지능  
(Artificial Intelligence), 자율주행 드론  
(Self-driving Drone) 등

### 김혜근(Hye-Geun Kim)



2019년~현재 : 전남대학교 전자컴퓨터  
공학부 소프트웨어공학  
전공 학사과정

※ 관심분야 : 컴퓨터 비전(Computer Vision), 인공지능  
(Artificial Intelligence) 등

### 윤경윤(Kyung-Yoon Yoon)



2019년~현재 : 전남대학교 AI융합대학  
IoT인공지능융합전공  
학사과정

※ 관심분야 : 컴퓨터 비전(Computer Vision), 인공지능  
(Artificial Intelligence) 등

### 황종원(Jong-Won Hwang)



2017년~현재 : 전남대학교 전자컴퓨터  
공학부 소프트웨어공학  
전공 학사과정

※ 관심분야 : 컴퓨터 비전(Computer Vision), 딥러닝(Deep  
Learning), 인공지능(Artificial Intelligence) 등

### 조영준(Yeong-Jun Cho)



2014년 : 광주과학기술원 정보통신공학  
(공학석사)  
2018년 : 광주과학기술원 전기전자컴퓨터  
공학 (공학박사-컴퓨터비전)

2018년~2020년: 현대모비스 데이터사이언스팀 책임

2020년~현재 : 전남대학교 소프트웨어공학과 조교수

※ 관심분야 : 컴퓨터비전(Computer Vision), 딥러닝(Deep  
Learning), 스마트팩토리 등