

3차원 실내 장면합성 응용을 위한 연구 현황 : 대화형, 최적화 및 딥러닝 방법 비교분석

조운식^{0,1} 이지원¹ 조세홍² 김진모^{1,2,*}

한성대학교 일반대학원 정보컴퓨터공학과¹, 한성대학교 컴퓨터공학부²
{yunsik.cho, mpm11, chosh, jinmo.kim}@hansung.ac.kr

Research Status for 3D Indoor Scene Synthesis Applications : A Comparative Analysis of Interactive, Optimization, and Deep Learning Methods

Yunsik Cho^{0,1} Jiwon Lee¹ Saehong Cho² Jinmo Kim^{1,2,*}

Department of Information and Computer Engineering, Graduate School, Hansung University¹

Division of Computer Engineering, Hansung University²

요 약

3차원 실내 장면합성 연구는 혼합현실, 로봇 협업, 디지털 트윈 등 다양한 산업과 응용 분야에서 중요한 연구 주제이다. 본 연구는 대화형 편집, 최적화 알고리즘, 그리고 딥러닝을 기반으로 하는 세 가지 핵심 연구 기술을 토대로 3차원 실내 장면합성 연구 현황을 특징, 장/단점, 한계 등으로 비교 분석한다. 대화형 편집 기반 연구들은 사용자 입력을 바탕으로 상황과 조건에 따라 의도를 반영한 객체 배치를 지원한다. 하지만 다양하고 자연스러운 장면합성결과를 위해서는 많은 시간과 노력이 필요하고 편집자의 숙련도에 영향을 받는다. 최적화 알고리즘 기반 연구들은 수학적 모델링과 비용 함수 등을 통해 높은 수준의 자동화를 구현하고 있다. 하지만, 긴 처리 시간과 지역 최소값 수렴 등으로 인한 부자연스러운 장면 연출에 대한 일반화된 해결방법이 모호하다. 마지막으로 딥러닝 기반 연구들은 대규모 데이터셋을 활용하여 높은 사실성과 다양한 연출이 가능하다는 특징이 있다. 하지만, 학습데이터의 의존성과 데이터 구축에 부담이 중요한 문제로 남는다. 본 연구는 기존의 연구 현황 분석을 통해 복합적 문제 해결 방향을 제안하고, 정량화된 모델 설계, 효율적 데이터 학습, 사용자 맞춤형 시스템 등 다양한 분야에서의 응용 방향을 논의한다.

Abstract

3D indoor scene synthesis research is an important research topic in various industries and applications such as mixed reality, robot collaboration, and digital twins. This study compares and analyzes the current status of 3D indoor scene synthesis research in terms of features, advantages/disadvantages, and limitations based on three core research technologies: interactive editing, optimization algorithms, and deep learning. Interactive editing-based studies support object arrangement that reflects intentions according to situations and conditions based on user input. However, diverse and natural scene synthesis results require a lot of time and effort and are affected by the skill of the editor. Optimization algorithm-based studies implement a high level of automation through mathematical modeling and cost functions. However, generalized solutions for unnatural scene production due to long processing times and local minimum convergence are ambiguous. Finally, deep learning-based studies are characterized by high realism and diverse productions by utilizing large-scale datasets. However, the dependency of learning data and the burden of data construction remain important issues. This study proposes a hybrid problem-solving direction through an analysis of the current state of research and discusses application directions in various fields such as quantitative model design, efficient data learning, and user-tailored systems.

키워드: 3차원 실내 장면합성, 대화형 편집, 최적화 알고리즘, 딥러닝

Keywords: 3D Indoor Scene Synthesis, Interactive Editing, Optimization Algorithms, Deep Learning

*corresponding author: Jinmo Kim / Hansung University (jinmo.kim@hansung.ac.kr)

1. 서론

목적과 상황에 부합하는 3차원 가상 장면은 가상환경을 경험하는 사용자의 몰입감을 높이고, 시각적 만족도에 직접적인 영향을 미친다. 이러한 이유로 효율적인 3차원 장면을 생성하거나 재구성하는 연구는 가상현실, 혼합현실 등 실감형 콘텐츠의 활용성을 높이는 방식으로 다양한 산업 분야에 응용하기 위한 필수적인 분야이다. 3차원 가상장면은 3DS Max, Maya, Blender와 같은 그래픽 소프트웨어를 활용하여 디자이너가 직접 제작하는 방법이 일반적이다. 또는, 제작된 그래픽 리소스를 활용하여 유니티, 언리얼과 같은 엔진을 통해 편집하는 방법이 있다. 하지만, 디자이너가 수작업으로 배치 및 편집하는 과정에서는 반복적인 작업이 많아지고, 디자이너의 편향이 반영될 수 있으며, 최종 결과물과 장면 제작 시간 또한 디자이너의 숙련도에 크게 좌우되는 문제가 있다. 이러한 이유로 실내 가상장면을 효과적이고, 자연스럽게 표현하기 위한 장면합성 연구가 진행되고 있다. 3차원 실내 장면합성 연구는 단순하고 반복적인 작업을 최소화하고, 합성 결과의 다양성을 높이기 위한 목적으로 다양한 접근의 연구들이 이루어지고 있다. 본 연구는 디자이너의 개입 정도에 따른 장면합성 연구의 장단점을 비교하기 위해 3차원 실내 장면합성 연구를 다음과 같은 세 가지 핵심기술로 나누어 정리한다.

1. **대화형 편집:** 객체추천 알고리즘을 기반으로 최소한의 사용자 입력을 통해 장면을 합성하는 방법
2. **최적화 알고리즘:** 비용함수 기반 최적해 알고리즘을 통해 장면합성을 자동화하는 방법
3. **딥러닝 응용:** 대규모 데이터셋을 활용하여 장면에 대한 규칙을 학습하고 이를 바탕으로 새로운 장면을 추론하는 방법

장면합성을 위해 실내 공간의 범주, 도면정보와 같은 레이아웃을 토대로 최소한의 사용자 입력을 토대로 배치할 객체를 추천하거나 자연스러운 장면합성 결과를 위한 최적의 위치, 방향 정보를 계산하는 대화형 편집 기반 연구들이 있다. Zhang et al. [1]은 3차원 장면합성 과정에서 마우스의 위치만으로 현재 공간에 배치될 추천 객체를 선별하는 알고리즘을 제안하였다. 대화형 편집 기술은 기존의 그래픽 소프트웨어를 통한 편집 과정을 줄여 빠르고 효과적으로 장면합성 결과를 유도함을 목적으로 하지만 작업자의 의존도는 여전히 활용에 중요한 문제로 남는다.

확률적 최적해 알고리즘, 샘플링 기법 등을 통한 수학적 모델을 활용한 장면합성 연구들은 주어진 표본 정보로부터 반복적인 계산을 통해 이상적인 결과를 도출하는 방법이다. 예를 들어, 마르코프 연쇄 몬테카를로 (Markov Chain

Monte Carlo) 기반의 메트로폴리스 헤이스팅스 (Metropolis-Hasting)와 같은 샘플링 기법, 담금질 기법 (Simulated Annealing)과 같은 전역 최적화 문제 대한 확률적 메타 알고리즘 등을 통해 주어진 표본 장면으로부터 실내 장면을 구성하는 객체의 배치 관계를 추출하고, 비용함수를 토대로 장면을 재구성하는 방법 등이 있다 [2]. Weiss et al. [3]은 물리학에 기반을 둔 연속 레이아웃 합성 기술을 제안하여 빠르고 정교한 장면합성 결과를 보여주기도 하였다. 하지만, 최적화 알고리즘을 기반으로 한 연구들은 복잡한 관계, 구조의 장면을 합성할 때 사전에 정의해야 할 요소, 비용함수의 수, 또는 매개변수가 많아져 일반화된 규칙을 설계하기가 어렵고, 이 과정에서 사용자의 편향이 반영될 수 있다는 한계가 있다. 이러한 이유로 Li et al. [4]은 비용함수 설계에 필요한 가중치를 사용자가 직접 설계하는 것이 아닌 다중 객체 입자 군집 최적화 (Multi Object Particle Swarm Optimization) 알고리즘을 활용하여 비지배 파레토 해집합을 구성하여 가중치를 직접 설계하지 않고도 다양한 장면합성 결과를 도출하기도 하였다.

마지막으로, 딥러닝 기반의 장면합성 연구들은 대규모의 3차원 실내 장면 데이터셋을 활용하여 장면을 구성하는 객체들의 규칙, 패턴을 암시적으로 학습하여 새로운 장면을 생성한다. 대표적으로, ATISS (Autoregressive Transformers for Indoor Scene Synthesis) [5]는 트랜스포머 기반 딥러닝 모델을 활용한 장면합성 연구로, 이후 다양한 딥러닝 기반 장면합성 연구의 기반이 된 중요한 선행 연구이다. 대규모의 3차원 실내 장면 데이터셋인 3D-FRONT [6]를 활용하여 각 장면의 레이아웃 구조와 그에 따른 객체 간의 관계를 자동 회귀 트랜스포머를 기반으로 학습하여 다양한 장면합성 결과를 도출한다. 이를 토대로 사용자의 상호작용을 고려한 장면합성 [7], 언어를 기반으로 3차원 장면을 생성하는 연구 [8], 명시적이지 않은 불완전한 장면을 온전하게 재구성하는 연구[9]들도 진행되었다. 이러한 딥러닝 기반의 장면합성 연구는 복잡한 장면의 패턴과 구조를 학습하고, 다양한 객체로 구성된 실내 장면을 실시간으로 자연스러운 결과를 도출할 수 있다는 점에서 강력한 도구로 자리 잡고 있지만, 많은 양의 학습데이터가 필요함은 물론 이에 의존된 결과와 함께 데이터셋의 품질과 크기가 결과에 미치는 영향이 크다는 한계가 있다. 특히, SUNCG [10]와 같은 대규모 데이터셋은 더는 사용이 불가능해지면서 SUNCG 데이터셋을 활용한 기존의 연구와 비교를 수행할 수 없다. 이로 인하여, 3D-FRONT와 같은 다른 3차원 실내 장면 데이터셋을 활용하여 재학습하고 비교실험을 수행해야 한다는 번거로움이 발생하기도 하였다.

최근 3차원 장면합성에 대한 다양한 조사가 진행되었다. Munoz-Silva et al. [11]은 3차원 장면 재구성을 위하여 이미지에서 포인트 클라우드를 생성하는 방법과 관련된 연구를 조사하였고, Bae et al. [12]은 그래프 기반 3차원 장

면 생성 학습모델에 관하여 조사하였다. 최근에는 딥러닝 기술의 발전과 더불어 딥러닝 기반 장면합성 연구들이 제시하는 Data, Coverage, Capture, Model Textures, 3d Annotation의 요소에 대해 비교분석을 수행하기도 하였다 [13]. 하지만 이러한 연구들은 3차원 장면합성의 특정 방법에 대한 조사를 수행하였을 뿐 다양한 실내 장면합성 연구의 비교분석을 수행하지 않아 각 장면합성 기술의 장단점을 전반적으로 다루지 않았다.

따라서, 본 연구는 현재까지 진행되고 있는 3차원 실내 장면합성 연구를 대화형 편집, 최적화 알고리즘, 그리고 딥러닝 기반의 세 가지 주요 범주로 분류하고, 각 범주 내에서의 연구 현황을 분석한다. 이를 통해 각 방법론의 기술적 장단점을 비교 평가하고 한계를 파악하며, 향후 3차원 실내 장면합성 연구가 나아가야 할 방향을 제시하고자 한다. 또한, 이러한 분석을 바탕으로 다양한 방법론 간의 통합 가능성과 새로운 연구 과제를 도출하여, 자동화된 3차원 실내 장면합성 기술의 발전에 기여하고자 한다.

2. 실내 장면합성 구조 및 개념

2.1 장면합성을 위한 고려사항

3차원 실내 장면을 제작하기 위하여 디자이너가 유니티 엔진이나 언리얼 엔진과 같은 게임엔진을 기반으로 그래픽 리소스 객체를 3차원 장면에 직접 배치, 편집하는 방식으로 진행된다. 3차원 실내 장면합성 연구는 디자이너의 수고와 편향을 줄이면서도 미적 기준과 상호작용성을 고려한 객체 배치를 통해 빠르고 다양한 장면합성을 목표로 한다. 이를 위해 3차원 실내 장면의 정보를 컴퓨터가 이해할 수 있는 데이터로 정의하는 실내 장면의 표현 과정이 필요하다. 실내 장면을 표현하기 위해서 공간의 전체적인 레이아웃, 장면을 구성하는 객체의 3차원 변환정보 (위치, 회전, 크기), 객체 레이블 정보 등을 기반으로 실내 장면을 정의한다. 또한, 객체와 객체 사이의 상관관계, 사용자 상호작용을 위한 여유 공간 등에 대한 정보들은 명시적으로 표현하고, 딥러닝 기반 실내 장면합성에서는 객체의 관계를 암시적으로 학습한다. 실내 장면의 레이아웃은 배치하고자 하는 장면의 바닥 정보나 실내 장면을 구성하는 벽 객체로 입력 레이아웃은 잠재적인 다양한 객체 배치 정보를 포함할 수 있다. 명시적인 실내 장면 표현 방법에서 객체는 주어진 레이아웃이 가지고 있는 벽 객체와의 거리, 회전각을 고려하여 배치하는 것이 일반적이다 (벽에 붙어있는 액자나 장롱 등). 따라서, 객체의 위치, 회전 정보를 기반으로 벽과의 거리, 회전 각도를 명시적으로 설계하는 과정이 필요하다[2]. 이와 더불어 객체 간 겹치지 않도록 배치하거나 장면 레이아웃에 벗어나는 형태의 객체 배치를 피하기 위해 장면을 구성하는

객체의 경계상자를 활용한다 [14, 15]. Figure 1은 3차원 실내 장면합성에 필요한 객체 표현의 예로, Yu et al. [2]의 논문에 제시된 그림이다.

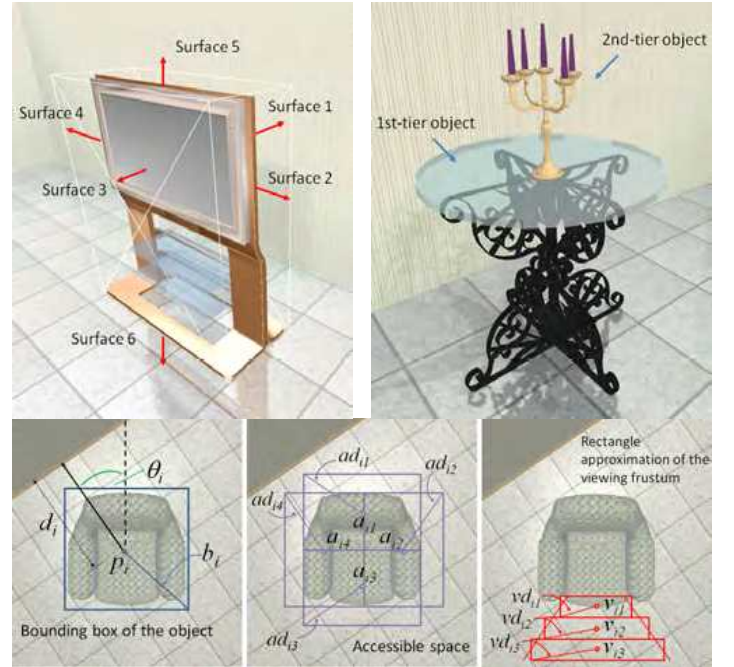


Figure 1. Example of an object representation method that constitutes an indoor scene [2].

2.2 데이터셋 현황

3차원 실내 장면 정보를 명시적으로 설계하는 것은 추가적인 수작업이 필요할뿐더러 설계과정에서 디자이너의 편향이 적용되어 일반화된 장면 표현에 대한 데이터를 구축하는데 제약이 될 수 있다. 이러한 이유로, 대규모 실내 장면 정보 데이터셋을 토대로 딥러닝 기반의 학습모델 훈련을 통해 실내 장면 데이터의 규칙을 파악하고, 이를 활용하여 자동화된 3차원 실내 장면합성을 수행할 수 있다. 모델 학습을 위한 3차원 장면 데이터셋은 각각 구성하고 있는 정보가 데이터셋의 목적에 따라 다르므로 데이터셋 각각의 구성 정보를 파악하여 활용하여야 한다. Table 1은 딥러닝 기반 장면합성에 필요한 데이터셋에 대한 이전 연구에 대한 분석으로 각 데이터셋의 특징을 정리하여 나타낸 것이다.

DataSet	Release	Models	Categories
ShapeNet [16]	2015	220,000	55
Thing10k [17]	2016	10,000	72
SUNCG [10]	2017	45,622	80
ScanNet [18]	2017	2,500,000	20
3D-FUTURE [19]	2020	9,992	34

Table 1. Information on 3D indoor scene datasets.

ShapeNet, Thing10k, SUNCG, ScanNet, 3D-FUTURE 데이터셋은 모두 3차원 장면 및 객체 정보를 담고 있는 중

요한 리소스로, 각각의 데이터셋은 다양한 형식의 메타데이터 파일 (JSON, STL 형식 등)을 제공하며, 이를 통해 객체 고유 식별자, 카테고리, 3차원 변환 정보(위치, 회전, 크기), 재질 및 텍스처, 경계상자, 객체 간의 관계 등을 포괄적으로 설명한다. ShapeNet은 3차원 CAD 모델 기반의 데이터셋으로, 220,000개의 모델이 WordNet 텍소노미를 기준으로 분류되어 있으며, 각 객체의 기하학적 속성과 시맨틱 정보를 제공하여 객체 인식 및 생성 연구에 활용된다. Thingi10k는 10,000개의 3차원 프린팅 모델로 구성된 데이터셋으로, Thingiverse 플랫폼에서 수집된 고품질 모델들이 포함되어 있다. STL 형식으로 제공되며, 모델의 메쉬 복잡도, 다각형 구조, 프린팅 가능성 등의 정보를 포함하여, 3D 프린팅 알고리즘 연구에 적합하다. SUNCG는 45,622개의 3차원 실내 장면으로 구성된 데이터셋으로, 각 장면은 개별적으로 라벨링된 3차원 객체 메쉬로 구성되어 있다. 이 데이터셋은 객체 범주와 함께 장면의 구조적 배치 및 시맨틱 정보를 포함하며, 3차원 장면합성 및 시맨틱 분할 연구에 활용된다. ScanNet은 실제 실내 환경을 기반으로 한 RGB-D 비디오 데이터셋으로, 2.5M 프레임의 데이터를 포함하며, 3D 표면 재구성, 카메라 포즈, 시맨틱 분할 정보가 포함되어 있다. 데이터셋 내 객체 모델 수는 약 2,500,000개에 달하지만, 이는 동일 객체가 여러 프레임에서 다양한 시점으로 중복 캡처된 결과이다. 이 데이터셋은 클라우드싱 방식으로 라벨링된 객체 정보와 CAD 모델 정렬 데이터를 제공하며, 3차원 장면 이해 및 객체 분류 연구에 유용하다. 3D-FUTURE는 3차원 가구 모델과 실내 장면을 다룬 대규모 데이터셋으로, 20,240개의 실내 이미지와 9,992개의 가구 모델이 포함되어 있다. 이 데이터셋은 고품질 텍스처, 카테고리, 재질 속성, 그리고 실내 배치 정보를 제공하며, 실내 디자인 및 텍스처 복원 연구를 위한 리소스를 제공한다. Figure 2는 딥러닝 기반 장면합성 연구에서 학습 모델을 훈련하기 위한 3차원 실내 장면 데이터셋을 나타낸 것으로, Fu et al. [19]의 연구 결과로 제시된 그림이다. 현재 SUNCG 데이터셋은 사용이 제한되어 있어, 딥러닝 기반 장면합성 연구에서 데이터셋 의존성이 문제로 지적되었다. 연구자들은 3D-FRONT 또는 3D-FUTURE와 같은 대체 데이터셋을 사용해 기존 연구를 재현하거나 새로운 실험을 설계하고 있지만, 이는 데이터셋에 의존도가 높은 만큼 딥러닝 장면합성 모델의 유지 보수 측면에서 데이터셋의 가용성, 품질, 확장성에 따라 연구의 지속 가능성이 크게 영향을 받을 수 있다는 한계를 내포하고 있다. 데이터셋이 더는 지원되지 않거나 제한적으로 제공될 경우, 새로운 데이터셋에 대한 재학습 및 모델의 재구성이 필요하며, 이는 추가적인 시간과 자원을 요구하게 될 수 있다.

3. 장면합성 연구 현황

3.1 대화형 편집

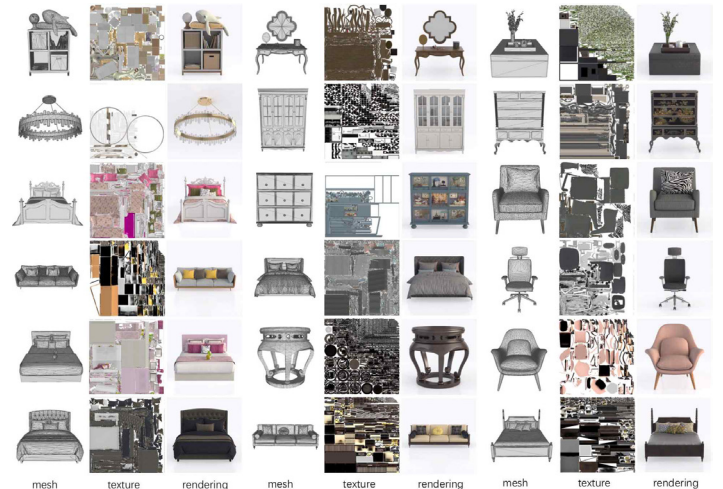


Figure 2. Example of 3D indoor scene data [19].

실내 장면 제작을 위해 디자이너가 수작업을 통해 진행하는 과정에서의 문제나 불편함을 해소하기 위하여 장면합성을 위한 저작도구 방식의 대화형 편집 기반의 연구들이 진행되었다. Figure 3은 Zhang et al. [1]의 MageAdd 프레임워크로 사용자가 마우스 커서를 움직이는 위치에 따라 적절한 객체를 실시간으로 추천하고, 자동으로 배치하는 인터페이스를 제공한다. 디자이너의 숙련도에 의존하지 않으며 사용자가 최소한의 입력으로 직관적이고 효율적으로 장면을 구성할 수 있도록 설계되었다. 그러나 개별 객체 배치에 초점을 맞추고 있어 복잡한 장면 구성에서 객체 간의 상호작용성이나 공간에서의 행동, 활동성을 고려하는 데 제약이 있다. You et al. [20]은 가상현실을 활용하여 사용자가 장면 전체를 탐색하고 재구성할 수 있는 환경을 제공하였다. 현실 공간을 스캔한 후, 가구 배치 및 스타일 변경을 직관적으로 시도할 수 있도록 지원하는 프레임워크를 제안한다. 그러나 가상현실 기반의 상호작용은 사용자에게 몰입감은 제공하지만, 가구 간의 세부적인 관계를 모델링하거나 대규모 장면에서 효율적인 작업을 지원하는 데는 한계가 있다. Zhang et al. [21]은 개별 객체가 아닌 그룹 단위의 객체 편집을 지원하여, 대규모 장면에서의 효율성을 크게 향상시켰다. 사용자가 특정 객체를 선택하면 관련된 객체 그룹 전체를 재배치하거나 수정할 수 있는 기능을 제공하여, 장면 내 객체 간의 상호작용을 효과적으로 관리할 수 있다. 하지만, 사용자가 전체 장면의 공간적 제약이나 레이아웃에 대한 명시적인 피드백을 제공하지 않는다는 단점이 있다. 이러한 공간적 제약 문제를 보완하기 위해 방 크기와 레이아웃의 변경에 따라 객체를 실시간으로 재구성하고, 공간 변화를 반영한 최적의 배치를 자동으로 제안하여 사용자 작업의 효율성을 크게 향상시켰다 [22]. 이처럼 대화형 편집 기반 장면합성 연구는 사용자의 편의를 높이면서 효과적으로

장면합성을 수행할 수 있도록 발전하였지만, 여전히 객체 배치의 다양성 부족, 사용자 의도나 목적의 반영 한계, 대규모 데이터 학습의 필요성 등 여러 과제가 남아 있다. 이러한 문제를 해결하려면 사용자와 시스템 간의 상호작용을 강화하고, 데이터 기반 학습을 효과적으로 통합한 새로운 접근법이 필요하다. Figure 3은 MageAdd [1] 와 SceneExpander [22]의 대화형 편집 저작도구 방식의 연구 결과 그림을 가져온 것이다.

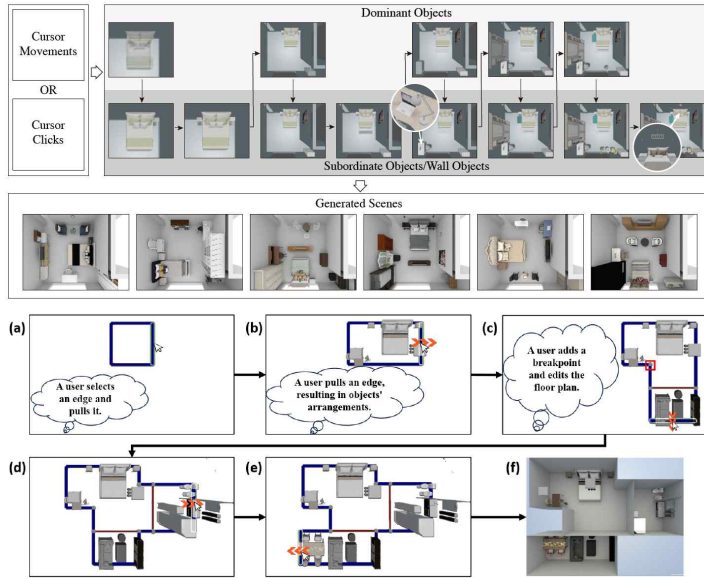


Figure 3. Process of scene synthesis based on interactive editing [1, 22].

3.2 최적화 알고리즘

최적화 알고리즘 기반의 장면합성 연구는 장면을 구성하는 객체의 배치를 위하여 공간에서의 특징, 객체 간의 관계, 객체 분류 등을 수학적으로 정의하고 자연스러운 장면합성 결과라는 최적의 해를 찾아가는 방법이다. 이와 관련된 연구 방법론은 주로 비용 함수와 최적화 알고리즘을 통해 주어진 해를 찾아가는 구조로 반복적인 연산을 통해 수행된다. Yu et al. [2]은 담금질 기법을 통해 최적의 해를 찾는 과정을 반복하며, 메트로폴리스 헤이스팅스 상태 탐색을 사용하여 레이아웃의 결과가 지역 최솟값 (local minima)에 수렴하지 않고 전역 최솟값 (global minima)을 탐색할 가능성을 검증하였다. 이상적인 표본 장면이 주어지면 이로부터 장면합성에 필요한 정보를 추출하고, 여기에 가시성, 이동 경로, 접근성과 같은 인간 공학적 요소나 심미적 요소를 포함한 제약 조건을 종합한 비용 함수를 정의한다. 그리고 샘플링 기법을 통해 초기에 무작위로 배치된 객체들을 점진적으로 조정해 나감으로써 자연스러운 장면합성 결과를 달성하는 전역 최솟값으로 수렴해갈 수 있도록 한다. Figure 4는 최적화 알고리즘 기반 장면합성의 전반적인 흐름을 보인다. 그러나 높은 계산 복잡성과 긴 처리 시간은 객체의 수가 많은

대규모 장면이나 상호작용성을 고려한 가상환경에 적용하기 어려운 한계를 보인다.

Kán and Kaufmann [23]은 유전 알고리즘 (Genetic Algorithm)을 사용하여 다차원 공간에서 객체 배치 문제를 해결하였다. 사용자가 객체를 미리 지정하지 않아도 되는 자동 객체 선택을 지원하고, 재질 (material) 선택과 색상 조화를 추가로 최적화 과정에 설계하여 결과물의 심미적 품질을 높인다. 그러나, 유전 알고리즘이 가지는 탐색 과정은 장면합성 결과 생성까지 시간적 문제는 개선의 여지가 있다. 탐욕적 비용 최소화 (Greedy Cost Minimization)를 활용한 연구는 유전 알고리즘 연구와 유사한 구조에서 빠르고 효율적인 가구 배치를 가능하게 한다 [24]. 이 방법은 전역 최적화를 목표로 하기보다는 지역적으로 최적화된 다수의 해를 생성하여 사용자의 다양한 요구를 충족시킨다. 또한, 절차적 방법을 활용하여 세부 객체를 효율적으로 배치하고, 실시간 상호작용이 가능한 속도를 제공하여 유전 알고리즘의 성능 한계를 극복하였다. 기존의 마르코프 연쇄 몬테카를로 확률적 최적화 기법과 비교하여 연속적 위치 기반 최적화 (Position-Based Optimization)를 도입하여 속도를 크게 향상한 연구가 수행되기도 하였다 [3]. 이는 물리적 타성 시뮬레이션의 공학적 개념을 객체 배치 문제에 적용하여 비선형 제약 조건을 연속적 절차로 해결한다. 이는 실내 장면은 물론 실외 장면 및 대규모, 밀집된 장면에서도 효과적인 결과를 나타냈지만, 제약 조건의 복잡성이 증가할수록 구현 난이도가 높아지는 한계가 있어 다양한 장면을 고려하는데 어려움이 따른다.

Liang et al. [25]는 개인화된 사용자 선호도를 학습하여 워크스페이스를 최적화하는 장면 재구성 연구를 수행하였다. 사용자가 가상현실 환경에서 작업을 수행하는 동안 활동 데이터를 수집하여 개인화된 비용 함수를 구성하고, 이를 기반으로 장면 레이아웃을 개선한다. 사용자 중심의 접근 방식은 혁신적이지만, 이 역시도 실시간 적용 가능성은 제한적이다. 최근에는 로봇 산업의 발전과 함께 인간과 로봇의 협업을 고려하여 실내 장면을 재구성하는 연구도 진행되고 있다 [26]. 적응형 담금질 기법 (Adaptive Simulated Annealing)과 공분산 행렬 진화 전략 (CMA-ES)을 기반으로 인간의 선호도와 로봇의 이동성을 동시에 고려하여 장면을 최적화하고, 로봇의 접근 가능성과 상호작용을 향상하는 새로운 연구 방향성을 제시하였다. 이처럼 최적화 알고리즘 기반의 장면합성 연구는 이상적인 장면합성 결과를 최적의 해로 정의하고 수학적 모델을 토대로 탐색, 계산 과정을 통해 사용자의 수작업과 사전 작업을 최소화 하여 자연스러운 실내 장면을 도출하는 방향으로 연구가 활발히 진행되고 있다. 하지만 계산 효율성과 제약 조건의 복잡성 사이의 균형을 맞추는 것이 주요 과제이며, 이는 결국 자연스럽고 다양한 장면합성 결과를 빠르고 효율적으로 생성하는 것과 직결될 수 있다는 과제로 남는다. Figure 4는 Yu et al. [2]과

Liang et al. [25]의 최적화 과정을 논문을 통해 제시한 그림을 가져온 것이다.

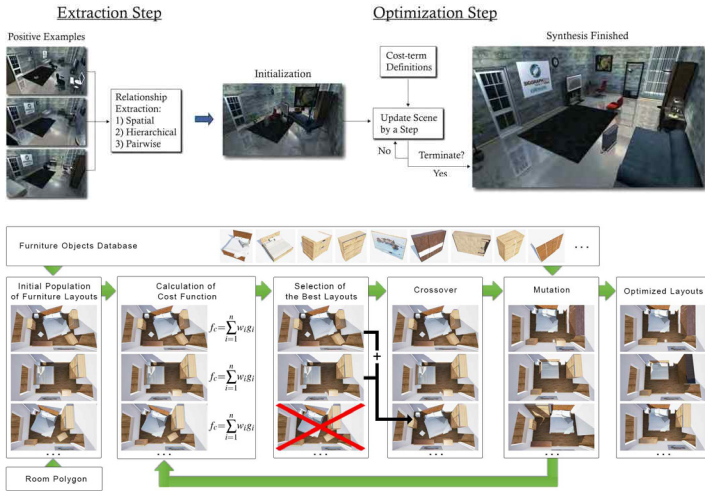


Figure 4. Examples of scene synthesis process using an optimization algorithm [2, 25]

3.3 딥러닝 응용

딥러닝 기반 장면합성 연구는 대규모 데이터셋과 심층 신경망을 활용하여 장면 구성을 위한 객체의 특성, 객체 간 관계, 공간적 규칙 등을 학습하고 새로운 실내 장면을 자동으로 생성하는 데 초점을 맞추고 있다. 최적화 알고리즘 기반 장면합성 연구에서 제약으로 남아 있는 복잡한 관계 또는 객체의 특성을 학습을 통해 해결하고 높은 수준의 사실성을 모델링 할 수 있는 점에서 중요한 진보를 이루었다.

Paschalidou et al. [5]은 자동 회귀 트랜스포머 (Autoregressive Transformer)를 활용하여 실내 장면을 순차적으로 생성하는 방식을 제안하였다. ATISS는 3D-FRONT와 같은 대규모 데이터셋에서 학습된 객체 배치 규칙을 기반으로 하며, 각 객체를 순차적으로 생성하면서 이전에 배치된 객체의 정보를 바탕으로 새로운 객체의 속성과 배치 위치를 예측한다.

이 방식은 객체 간의 관계를 세밀히 학습할 수 있는 강점을 가지지만, 순차적 생성 특성상 처리 속도가 느릴 수 있다는 한계도 있었다. 이러한 한계를 해결하기 위해 Tang et al. [27]은 Denoising Diffusion 모델을 사용하여 객체 집합을 순서에 의존하지 않는 방식으로 학습 및 생성하는 DiffuScene을 제안하였다. DiffuScene은 객체의 위치, 크기, 방향, 의미론적 속성을 포함한 장면을 생성하고, 텍스트 지시를 기반으로 한 조건부 생성을 지원한다. 이를 통해 다양한 장면을 물리적으로 타당하게 생성할 수 있고, 기존 모델 대비 더 높은 유연성과 사실성을 제공하였다. 또한, 사용자 정의 가능성을 강조한 연구로 인코더-디코더 아키텍처를 통해 객체의 속성(예: 위치, 크기, 방향)을 제어 가능한 방식으로 생성하는 연구가 수행되었다 [28]. 기존의 순차적

생성 방식에서 벗어나, 조건부 생성을 통해 사용자가 특정 객체의 속성을 고정하거나 조정할 수 있는 기능을 제공함으로써 디자인 과정의 자유도를 크게 확장하였다.

Min et al. [29]은 장면 구성의 기능성을 보장하기 위해 변형 가능한 자동 인코더 (Variational Autoencoder)를 활용한 연구도 수행되었다. 이 연구는 방의 경계 조건에 따라 기능 그룹을 생성하고, 각 그룹 내에서 객체를 배치함으로써 실내 장면의 기능성과 실용성을 충족시키고 다단계 제어를 가능하게 한다. Wei et al. [30]은 기존의 어수선한 장면을 정리된 상태로 변환하려는 연구로 Lego-net을 제안하였다. 이 모델은 Diffusion 모델을 기반으로 객체의 초기 상태를 분석하여 최소한의 이동으로 정리된 결과를 생성한다. 기존 장면의 분위기를 유지하면서도 규칙적이고 체계적인 배치를 가능하게 하는 이 접근법은 데이터 기반의 정리 알고리즘을 제시하였다. 더 나아가, 장면의 계층적 구조를 학습하는 연구도 수행되었는데 이는 변형 가능한 재귀 자동 인코더 (Variational Recursive Autoencoder)를 토대로 객체 간의 지지 관계, 인접성, 공존 관계를 학습하여 새로운 장면을 생성하며, 전반적인 구조와 세부적인 관계를 동시에 학습할 수 있도록 설계되었다 [31].

Sun et al.의 Haisor [32]는 인간-장면 상호작용을 고려하여 장면 배치를 최적화하는 연구로, 강화학습 (Reinforcement Learning)을 기반으로 한다. 이 모델은 Dueling Double DQN 및 몬테카를로 트리 탐색 (Monte Carlo Tree Search)을 활용하여 상호작용 공간을 확보하거나 넓은 여유 공간을 확보하는 등 인간 활동에 적합한 레이아웃을 생성하고, 가구 충돌을 최소화하여 사용자의 편의성을 극대화하는 방식으로 학습이 진행되었다. 이와 같은 목적으로, Zhao et al. [33]은 대규모 장면 생성을 지원하는 연구 LUMINOUS를 제안하였다. LUMINOUS는 Constrained Stochastic Scene Generation (CSSG) 알고리즘을 통해 합리적이고 실용적인 장면을 생성하고, Embodied AI 연구를 위한 다중 모드 데이터를 제공한다. 이는 대규모 데이터 생성 및 학습에 적합한 환경을 지원한다. 마지막으로, Proximal Policy Optimization 알고리즘을 사용하여 객체를 순차적으로 배치하며, 제약 조건과 보상 함수를 통해 장면 생성을 최적화하는 연구도 수행되었다 [34]. 이는 데이터셋에 대한 의존성을 줄이고, 다양한 장면을 효율적으로 생성할 수 있는 강점을 가진다. Figure 5는 딥러닝 기반 장면합성 연구의 예로, 자동 회귀 트랜스포머와 몬테카를로 트리 탐색을 통한 심층 Q-Learning 기반 장면합성 네트워크 모델 학습 구조를 보인다. Figure 5는 ATISS [5]와 HAISOR [32] 연구에서 제시한 학습 네트워크 모델의 그림을 가져온 것이다.

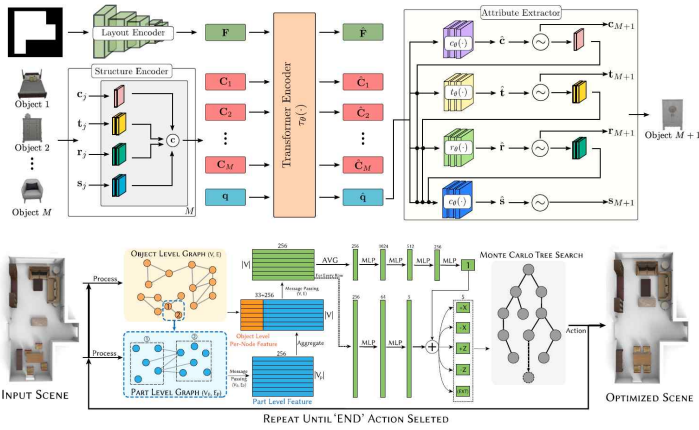


Figure 5. Scene synthesis training network based on autoregressive transformer and deep reinforcement learning [5, 32].

Table 2는 딥러닝 기반 장면합성의 최신 연구 [37]에서 정량적 실험에 대한 비교를 수행한 표를 가져온 것이다. Forest2seq [37]는 최근에 진행된 딥러닝 기반 장면합성 연구로, 학습모델의 정량적인 성능평가를 위한 도구를 활용하여 제안하는 장면합성모델의 성능을 검증하였다. 딥러닝 기반 모델의 성능평가 도구로 KL (Kullback-Leibler divergence) 발산은 두 확률분포의 차이를 측정하는 지표, FID (Fréchet Inception Distance)는 학습모델이 생성한 결과가 실제 데이터와 얼마나 유사한지 평가하는 지표, CAS (Classifier Accuracy Score)는 생성 모델의 레이아웃 적합도를 평가하는 지표이며 KL, FID는 값이 낮을수록, CAS 값은 50%에 가까울수록 좋은 성능의 모델임을 나타낸다. Figure 6은 Forest2seq [37]의 장면합성 결과와 기존 딥러닝 기반 장면합성 연구[27, 28]들과의 결과를 논문을 통해 제시한 그림으로, 불완전한 장면이 입력 (Layout, Partial scene)으로 주어졌을 때 장면합성 결과를 보이고, Figure 7은 ATISS [5]의 장면합성결과를 바탕으로 사용자 상호작용성을 고려한 심층 강화학습 기반 장면합성 연구 [32]로 과정을 논문에서 제시한 그림이다.

이처럼 딥러닝 기반 장면합성 연구는 기존에 명시적으로 제약조건을 정의하는 방법론의 한계를 극복하고, 학습을 통해 장면이 가지는 복잡한 특성과 관계를 효과적으로 파악하였으며 현실적이고 다양한 3차원 장면합성을 가능하게 하였다. Table 1과 같은 대규모 데이터셋과 결합하여 최적해 알고리즘으로는 표현하기 어려웠던 높은 수준의 사실적 표현과 자동화를 구현하고 있지만, 데이터셋에 대한 높은 의존도는 여전히 활용에 많은 제한이 되고 있다. 장면합성 연구에 사용되는 대표적인 데이터셋인 3D-FRONT에서도 거실, 침실, 식당 장면만을 제공하고 있어 사용자가 원하는 다양한 목적, 상황에 맞는 실내 장면에 바로 활용할 수 없는 문제가 있다.

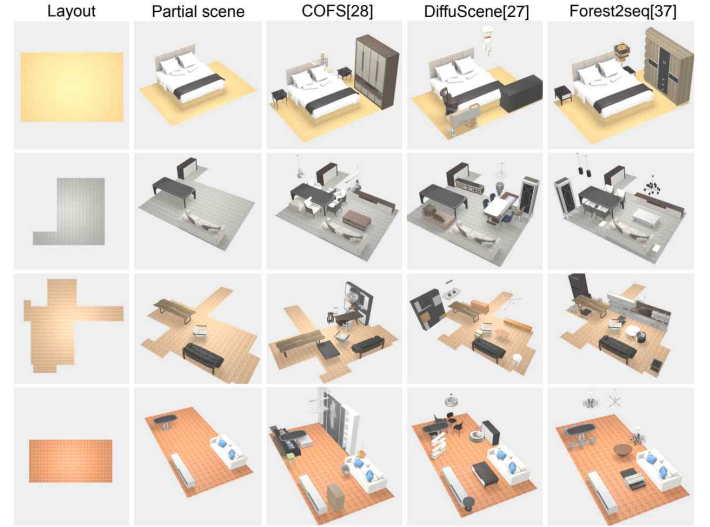


Figure 6. Example of deep learning-based scene synthesis results according to various inputs (Layout, Partial scene) [37].

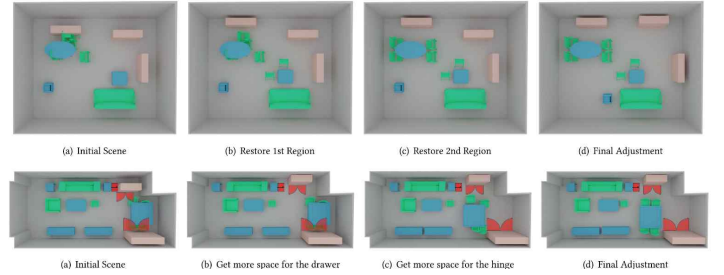


Figure 7. Process of performing scene synthesis considering user interaction using ATISS [5]-based scene synthesis results as input [32].

4. 토의

본 연구는 실내 장면합성 연구를 대화형 편집, 최적화 알고리즘 그리고 딥러닝 기반의 핵심기술로 분류하여 비교분석하고, 각 연구 방법론에 관한 장단점을 논의하기 위하여 연구 현황을 조사하였다. 대화형 편집 기반의 장면합성 연구는 디자이너의 직관적 입력을 통해 장면 구성의 효율과 의도를 직접 반영하는 데 중점을 둔다. 간단한 사용자 입력을 통해 현재 상태에서 가장 적절한 객체를 추천하고 배치하는 지원 방식으로 수작업의 반복성을 줄이는 데 이점을 보였지만 장면 구성이 복잡하거나 상황이나 연출이 다양해지면 디자이너의 숙련도에 많은 의존을 하게 되고 작업 효율도 감소하게 된다. 최적화 알고리즘 기반의 장면합성 연구는 수학적 모델링과 제약 조건에 기반을 둔 알고리즘 등을 토대로 장면합성에 필요한 객체 배치 과정을 자동으로 수행한다. 이와 관련하여, 담금질 기법, 유전 알고리즘, 탐욕 알고리즘, 물리적 탄성 시뮬레이션 등 다양한 방법론과 알고리즘을 통해 자연스러운 장면을 효과적으로 생성하거나 최적화의 정확도 또는 성능을 향상하기도 하였다.

Methods	Bedroom			Dining room			Living room			Library		
	KL↓	FID↓	CAS(%)	KL↓	FID↓	CAS(%)	KL↓	FID↓	CAS(%)	KL↓	FID↓	CAS(%)
FastSyn [35]	6.4	88.1	88.3	51.8	58.9	93.5	17.6	66.6	94.5	43.1	86.6	81.5
SceneFormer [36]	5.2	90.6	97.2	36.8	60.1	71.3	31.3	68.1	72.6	23.2	89.1	88.0
LayoutGPT [8]	17.5	68.1	60.6	—	—	—	14.0	76.3	94.5	—	—	—
ATISS [5]	8.6	73.0	61.1	15.6	47.6	69.1	14.1	43.3	76.4	10.1	75.3	61.7
COFS [28]	5.0	73.2	61.0	9.3	43.1	76.1	8.1	35.9	78.9	6.7	75.7	66.2
DiffuScene [27]	5.1	69.0	59.7	7.9	45.8	70.6	8.3	38.2	75.1	—	—	—
Forest2Seq [37]	4.2	67.9	58.3	5.5	40.2	65.6	5.9	35.2	68.0	5.2	69.1	57.3

Table 2. Quantitative comparative experimental results in deep learning-based scene synthesis research [37].

Items	Interactive Editing	Optimization algorithm	Deep-Learning
Methodology	User input-based, object recommendation and group editing	Definition of cost function, mathematical model such as global optimization, sampling, etc.	Training large datasets
Data dependency	Object model resources	Sample scene, user-defined parameters	Large-scale 3D scene dataset
Cost of calculation	Real-time interactivity	Complex computational complexity due to nonlinear optimization	High cost in learning process, relatively fast inference process
Degree of user input	High, user-dependent	Low, initial constraints and parameter settings	Usually, parameter setting during learning process
Diversity of results	Depends on user's capabilities	Depends on the design level of the constraints	Depends on the dataset
Strength	Intuitive, real-time editing capabilities	Setting explicit constraints	Automation, realism, diversity
Weakness	Highly dependent on user skill	High computational volume and difficulty in designing cost functions	Depends on dataset quality and availability, and has high training costs.

Table 3. Comparative analysis of interactive editing, optimization algorithms, and deep learning-based scene synthesis researches.

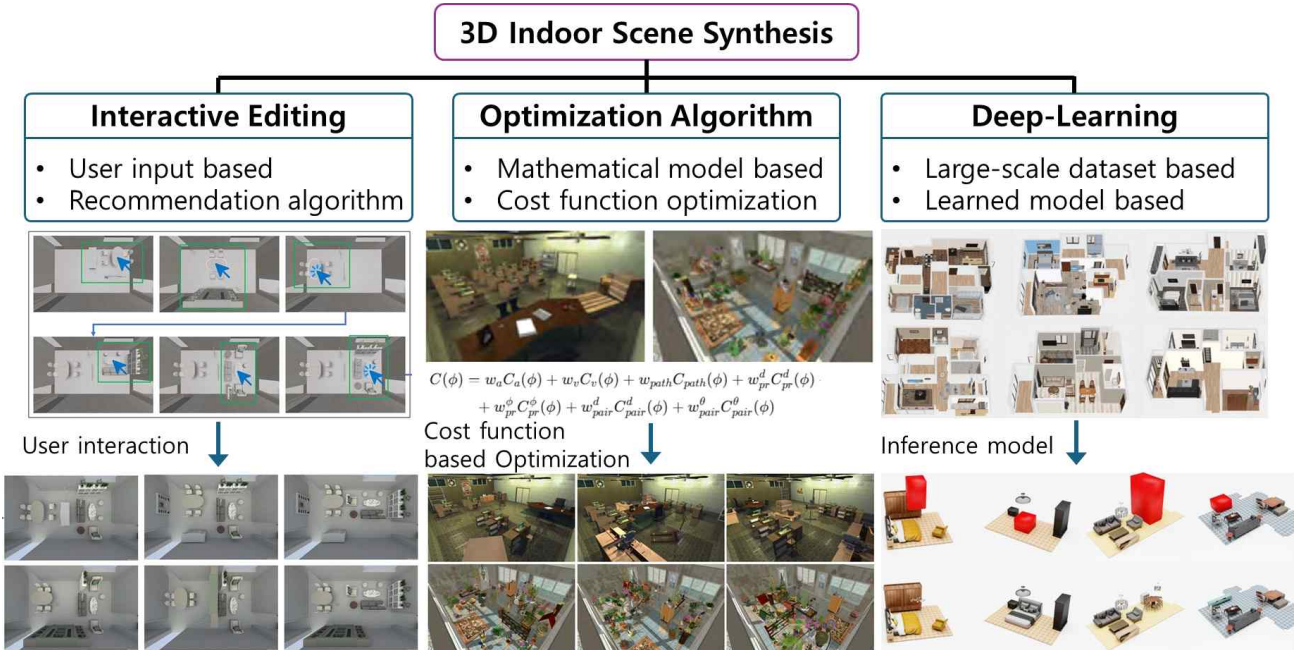


Figure 8. Comparison of scene synthesis processes based on interactive editing [21], optimization algorithms [2], and deep learning applications [5, 6]

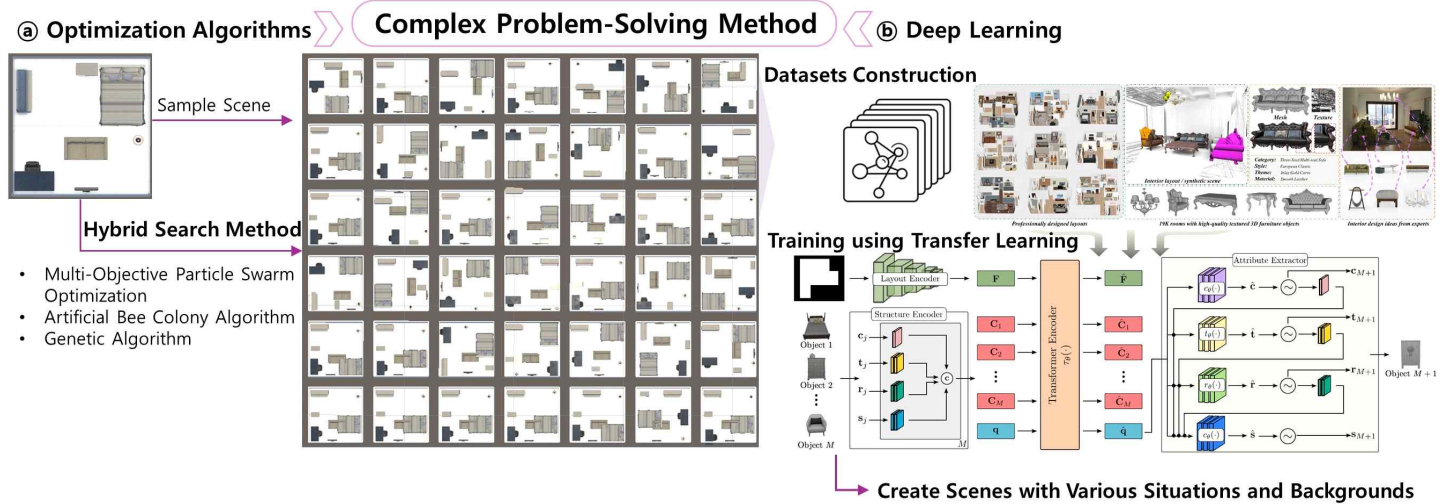


Figure 9. An example of a complex problem-solving method that uses an optimization algorithm based on a hybrid search method to generate various 3D scene synthesis results and use them as training data for deep learning-based scene synthesis research [5, 6].

Research Area	Goal	Technologies
Hybrid Method Development	Developing a scene synthesis methodology combining deep learning and optimization algorithms	Reinforcement Learning, Optimization Algorithms
Reducing Data Dependency Using Transfer Learning	Developing a model that can learn from existing datasets without requiring new data	Transfer Learning, Data Augmentation
Interactive Editing Tool Research	Researching interfaces that allow users to interactively modify scenes	UI/UX, Interaction Design, Recommendation algorithm
Optimization for Real Time Application	Enhancing real-time responsiveness in 3D content	Lightweight Algorithms, Real-time Rendering
User Preference Based Model Development	Developing a system that learns user feedback to automatically optimize scene layouts	Reinforcement Learning, Recommendation Systems
Exploring Applications in Various Fields	Exploring the potential applications of scene synthesis in VR, AR, and Robotics	Computer Vision, 3D Reconstruction, System Integration

Table 4. Future Research Directions in 3D Scene Synthesis.

그러나 실시간 상호작용이 가능한 정도의 빠른 장면합성 결과와 심미적 요인을 동시에 충족하는 결과를 모두 만족하는 데 한계가 있다. 마지막으로 딥러닝 기반의 장면합성 연구는 대규모 데이터셋에서 학습한 장면 정보를 토대로 레이아웃이 주어지면 자동으로 사실적인 결과를 빠르게 생성함을 보여주었다. 하지만, 데이터셋 품질과 가용성에 대한 의존도가 크다는 명확한 한계가 있다. 데이터셋에서 제공하지 못하는 상황의 장면을 표현하기 위해서는 데이터셋 구축부터 학습까지 일련의 과정을 모두 진행해야 하므로 응용 관점에서는 특히 어려움이 따를 수밖에 없다. Table 3은 대화형 편집, 최적화 알고리즘, 그리고 딥러닝 기반 장면합성 연구의 특징을 체계적으로 정리하여 비교한 내용이며 Figure 8은 각 방법의 장면합성 연구를 기반으로 입력부터 출력까지의 장면합성 과정을 도식화하여 비교하여 볼 수 있도록 기존의 논문을 통해 제시한 그림을 기반으로 한 그림이다.

5. 응용을 위한 연구 방향

3차원 실내 장면합성과 관련한 여러 연구는 각 핵심기술에 따라 강점과 약점을 가지고 있으며 이들의 결합을 통해 상호 보완적인 시스템을 제안할 수 있다. 예를 들어, 대화형 편집 기반의 방법과 최적화 알고리즘 기반 방법의 장점을 결합하여 최적해 탐색 과정에서 대화형 편집 기능을 추가하여 지역적 최솟값으로 수렴하는 과정을 예방하거나 빠르게 전역 최적해에 도달할 수 있도록 유도하는 방식으로 제안할 수 있다. 또한, 유전 알고리즘 또는 다중 목적 입자 군집 최적화 (Multi-Objective Particle Swarm Optimization) 알고리즘과 같은 여러 후보 해를 포함하는 방법을 활용하여 하나의 장면합성 결과가 아닌 다수의 결과를 동시에 생성하는 새로운 방법을 제안함으로써 다양한 장면 제작 결과를 효과적으로 생성함은 물론 딥러닝 기반 장면합성 연구에 필요한 데이터셋을 구축하는데 활용할 수 있을 것으로 기대한다. 이는 딥러닝 기반 장면합성 연구가 가지는 데이터셋 의존의 한계를 극복하는 새로운 연구 방향이 될 수 있을 것이며, 기존의 데이터셋이 제공하지 못하는 다양한 상황, 연출,

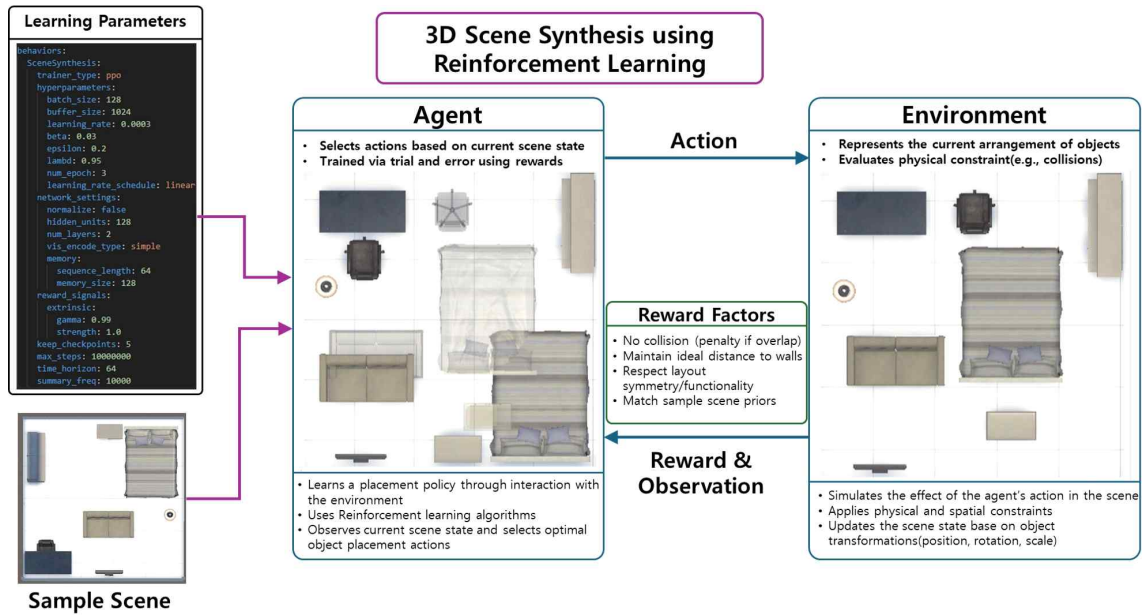


Figure 10. An example of how to generate 3D scene synthesis results based on reinforcement learning methods.

컨셉이 필요한 장면 생성이 필요한 경우 데이터셋을 빠르게 구축하고 딥러닝 기반 장면합성 연구를 응용하여 장면 생성에 활용할 수 있을 것이다.

현재의 3차원 실내 장면합성 연구는 데이터셋 의존성, 높은 연산 자원 요구, 사용자 맞춤화 부족 등의 한계를 가진다. 특히, 딥러닝 기반의 연구는 대규모 데이터셋의 품질과 가용성에 따라 성능이 좌우되고 딥러닝 기반 장면합성 연구에서 폭넓게 활용되던 SUNCG 데이터셋의 사용 제한 등의 문제들로 새로운 데이터셋을 재학습해야 하는 부담 등이 발생하고 있다. 또한, 최적화 알고리즘과 딥러닝 기반의 연구는 높은 계산 비용을 요구해 장면합성 결과만을 도출하는 것이 아닌 게임과 같은 3차원 대화형 콘텐츠나 가상현실 등 실감 미디어 분야에서 활용이 가능한 형태로 제공하는 데 필요한 실시간 적용, 사용자 의도 반영, 상호작용성 지원 등의 문제들이 여전히 남아 있다. 본 연구는 이러한 문제를 해결하기 위한 향후 연구 방향을 제시한다(Table 4). Figure 9은 본 연구에서 제시하는 복합적 문제 해결 방안의 하나로, 복합탐색법 기반의 최적화 알고리즘을 활용하여 하나의 최적 해가 아닌 다수의 다양한 장면합성 결과를 효과적으로 생성하고, 이를 ATISS [5]와 같은 딥러닝 기반 장면합성 연구의 데이터셋으로 적용하여 기존의 데이터셋 [6]이 제공하지 못하는 새로운 장면을 생성하는 과정을 보여준다. 대화형 편집, 최적화 알고리즘, 그리고 딥러닝 기반의 연구 방법을 결합한 하이브리드 방법론에 관한 계산 복잡성을 줄이고 실시간 적용성을 높인 포괄적인 연구가 필요하다. 또한, 다양한 장면합성 결과를 고려한 데이터셋 구축이 필요하고 전이학습 (Transfer Learning) 등과 같은 효율적인 학습 기법을 통해 데이터 의존성을 줄여야 한다. 현재까지는 대규모 데이터셋을 활용한 지도학습 (Supervised

Learning) 기반의 딥러닝 장면합성 연구가 주를 이루고 있으나, 향후에는 강화학습을 기반으로 장면 구성 규칙을 학습하는 방식의 연구도 주목할 필요가 있다. 강화학습을 통해 장면 내 객체 배치를 제어하는 에이전트를 설계하고, 비용 또는 보상 함수를 정의하여 반복적인 상호작용을 통해 스스로 최적의 배치 전략을 학습하도록 유도할 수 있다. 이는 데이터에 대한 의존도를 줄이고, 다양한 환경 변화에 적응할 수 있는 장면합성 방법을 제시한다. Figure 10은 강화학습을 활용한 장면합성 방식의 구조적 개요를 제시하며, 해당 접근 방식의 핵심 요소들을 요약하고 있다. 최종적으로 장면합성 자동화의 향후 연구 방향으로 사용자 선호를 학습하고 반영할 수 있는 대화형 딥러닝 모델을 구축하고, 가상현실, 증강현실, 로봇 협업 등 다양한 분야에서 장면합성 기술의 응용 가능성을 탐구할 필요가 있다.

6. 결론

본 연구는 3차원 실내 장면합성 응용을 위한 연구 현황을 비교 분석함을 목적으로 대화형 편집, 최적화 알고리즘, 그리고 딥러닝 응용의 세 가지 핵심기술로 연구를 분류하여 상세히 조사하였다. 각 연구 방법에 따른 장단점, 특징, 그리고 기술적 한계를 분석하였다. 대화형 편집 기반 장면합성 연구는 디자이너의 직관적인 입력을 바탕으로 장면 생성에 필요한 정보 (배치, 종류 등)를 효과적으로 지원하였으나 복잡한 구조나 객체 간 상호작용을 충분히 고려하지 못하는 문제를 인식하였다. 최적화 알고리즘 기반의 장면합성 연구는 최적해 탐색 알고리즘, 샘플링 기법을 비롯한 다양한 수학적 모델을 토대로 자동으로 자연스러운 장면합성 결과를 도출함을 확인하였지만, 계산 복잡성과 연산 자원 요

구로 인하여 실시간 응용에 제한이 있음을 확인하였다. 마지막으로 딥러닝 기반 장면합성 연구는 대규모 데이터셋을 활용하여 장면을 구성하는 객체의 특징, 복잡한 관계 등을 학습하고 높은 수준의 사실성을 구현하였으나, 딥러닝 연구가 가지는 데이터셋 의존성과 모델 재학습 부담이 응용에 부담이 되는 주요 문제로 지적되었다.

각 핵심기술을 기반으로 하는 연구의 고유한 장점을 결합한 복합적 문제 해결 방법은 실내 장면합성 연구를 다양한 분야에 응용하는데 중요한 방향성을 제시할 수 있다. 예를 들어, 최적화 알고리즘의 반복적 연산 과정에서 정확도 향상, 빠른 전역 최적 해로의 수렴 또는 사용자 의도 반영 등에 대화형 편집 방법을 적용하거나 딥러닝 기반 연구를 활용하여 학습데이터를 적절히 수학적 모델에 반영할 수 있다. 또한, 최적화 알고리즘을 변형하여 하나가 아닌 다수의 장면합성 결과를 정확하고 효과적으로 생성하는 연구를 딥러닝 기반 연구의 데이터셋으로 응용할 수 있는 프레임워크를 설계하여 데이터셋에 의존된 제한된 장면합성 결과만을 생성할 수 있게 하였던 기존 연구의 한계를 극복하고 다양한 상황, 연출이 반영된 장면합성에 딥러닝 기반 연구들을 활용함으로써 다양한 산업 분야로의 응용이 가능한 방향을 제시할 수 있을 것이다. 향후 다분야 응용 가능성을 고려하여 가상현실, 증강현실, 로봇 협업 등 다양한 환경에서 활용할 수 있는 장면합성 기술을 개발하는 데 초점을 맞추어야 한다. 이러한 연구는 3차원 장면합성 기술의 정확성과 효율성을 향상시키고, 다양한 응용 분야에서의 실질적인 활용을 가능하게 할 것이다.

감사의 글

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(연구개발 과제번호: 2021-0-00884, 비대면 협업용 솔루션 및 블록체인 기반 디지털 워크 통합 플랫폼 개발).

References

- [1] Zhang, S. K., Li, Y. X., He, Y., Yang, Y. L., & Zhang, S. H., "Mageadd: Real-time interaction simulation for scene synthesis," *In Proceedings of the 29th ACM International Conference on Multimedia*, pp. 965-973, 2021.
- [2] Yu, L. F., Yeung, S. K., Tang, C. K., Terzopoulos, D., Chan, T. F., & Osher, S. J., "Make it home: automatic optimization of furniture arrangement," *ACM Transactions on Graphics (TOG)-Proceedings of ACM SIGGRAPH 2011*, v. 30,(4), July 2011.
- [3] Weiss, T., Litteneker, A., Duncan, N., Nakada, M., Jiang, C., Yu, L. F., & Terzopoulos, D., "Fast and scalable position-based layout synthesis," *IEEE Transactions on Visualization and Computer Graphics*, 25(12), 3231-3243, 2018.

- [4] Li, Y., Wang, X., Wu, Z., Li, G., Liu, S., & Zhou, M., "Flexible indoor scene synthesis based on multi-object particle swarm intelligence optimization and user intentions with 3D gesture," *Computers & Graphics*, 93, 1-12, 2020.
- [5] Paschalidou, D., Kar, A., Shugrina, M., Kreis, K., Geiger, A., & Fidler, S., "Atiss: Autoregressive transformers for indoor scene synthesis," *Advances in Neural Information Processing Systems*, 34, 12013-12026, 2021.
- [6] H. Fu et al., "3d-front: 3d furnished rooms with layouts and semantics," *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10933-10942, 2021.
- [7] Yi, H., Huang, C. H. P., Tripathi, S., Hering, L., Thies, J., & Black, M. J., "MIME: Human-aware 3D scene generation," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12965-12976, 2023.
- [8] W. Feng et al., "Layoutgpt: Compositional visual planning and generation with large language models," *Advances in Neural Information Processing Systems*, 36, 2024.
- [9] Wei, Q. A., Ding, S., Park, J. J., Sajani, R., Poulenard, A., Sridhar, S., & Guibas, L., "Lego-net: Learning regular rearrangements of objects in rooms," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19037-19047, 2023.
- [10] Song, S., Yu, F., Zeng, A., Chang, A. X., Savva, M., & Funkhouser, T., "Semantic scene completion from a single depth image," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1746-1754, 2017.
- [11] Munoz-Silva, E. M., González-Murillo, G., Antonio-Cruz, M., Vázquez-Gómez, J. I., & Merlo-Zapata, C. A., "A Survey on Point Cloud Generation for 3D Scene Reconstruction," *In 2021 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE)*, pp. 82-87, IEEE, 2021.
- [12] Bae, J., Shin, D., Ko, K., Lee, J., & Kim, U. H., "A survey on 3d scene graphs: Definition, generation and application," *In International Conference on Robot Intelligence Technology and Applications*, pp. 136-147, 2022.
- [13] Patil, A. G., Patil, S. G., Li, M., Fisher, M., Savva, M., & Zhang, H., "Advances in DataDriven Analysis and Synthesis of 3D Indoor Scenes," *In Computer Graphics Forum* (Vol. 43, No. 1, p. e14927), 2024.
- [14] Germer, T., & Schwarz, M., "Procedural Arrangement of Furniture for Real-Time Walkthroughs," *In Computer Graphics Forum* (Vol. 28, No. 8), pp. 2068-2078, 2009.

- [15] Kjølås, K. A. H., “Automatic furniture population of large architectural models,” (*Doctoral dissertation, Massachusetts Institute of Technology*), 2000.
- [16] A. Chang et al., “Shapenet: An information-rich 3d model repository,” *arXiv preprint* arXiv:1512.03012, 2015.
- [17] Zhou, Q., & Jacobson, A., “Thing10k: A dataset of 10,000 3d-printing models,” *arXiv preprint* arXiv:1605.04797, 2016.
- [18] Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., & Nießner, M., “ScanNet: Richly-annotated 3d reconstructions of indoor scenes,” *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5828-5839, 2017.
- [19] Fu, H., Jia, R., Gao, L., Gong, M., Zhao, B., Maybank, S., & Tao, D., “3d-future: 3d furniture shape with texture,” *International Journal of Computer Vision*, 129, 3313-3337, 2021.
- [20] You, W., Shao, Y., Zheng, Z., Lu, Y., Yang, C., Zhou, Z., & Sun, L., “Old house new home: Facilitating interior design with RedesignUS in virtual reality,” *Displays*, 80, 102555, 2023.
- [21] Zhang, S. K., Tam, H., Li, Y., Ren, K. X., Fu, H., & Zhang, S. H., “Scenedirector: Interactive scene synthesis by simultaneously editing multiple objects in real-time,” *IEEE Transactions on Visualization and Computer Graphics*, 2023.
- [22] Zhang, S. K., Huang, J., Yue, L., Zhang, J. T., Liu, J. H., Lai, Y. K., & Zhang, S. H., “SceneExpander: Real-time scene synthesis for interactive floor plan editing,” *In Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 6232-6240, 2024.
- [23] Kán, P., & Kaufmann, H., “Automated interior design using a genetic algorithm,” *In Proceedings of the 23rd ACM symposium on virtual reality software and technology*, pp. 1-10, 2017.
- [24] Kán, P., & Kaufmann, H., “Automatic furniture arrangement using greedy cost minimization,” *In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 491-498. IEEE, 2018.
- [25] Liang, W., Liu, J., Lang, Y., Ning, B., & Yu, L. F., “Functional workspace optimization via learning personal preferences from virtual experiences,” *IEEE transactions on visualization and computer graphics*, 25(5), 1836-1845, 2019.
- [26] Wang, W., Zhao, Z., Jiao, Z., Zhu, Y., Zhu, S. C., & Liu, H., “Rearrange indoor scenes for human-robot co-activity,” *In 2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11943-11949, IEEE, 2023.
- [27] Tang, J., Nie, Y., Markhasin, L., Dai, A., Thies, J., & Nießner, M., “Diffuscene: Denoising diffusion models for generative indoor scene synthesis,” *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20507-20518, 2024.
- [28] Para, W. R., Guerrero, P., Mitra, N., & Wonka, P., “COFS: Controllable furniture layout synthesis,” *In ACM SIGGRAPH 2023 Conference Proceedings*, pp. 1-11, 2023.
- [29] Min, W., Wu, W., Zhang, G., & Zheng, L., “FuncScene: Function-centric indoor scene synthesis via a variational autoencoder framework,” *Computer Aided Geometric Design*, 111, 102319, 2024.
- [30] Wei, Q. A., Ding, S., Park, J. J., Sajjani, R., Poulenard, A., Sridhar, S., & Guibas, L., “Lego-net: Learning regular rearrangements of objects in rooms,” *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19037-19047, 2023.
- [31] Li, M., Patil, A. G., Xu, K., Chaudhuri, S., Khan, O., Sh Amir, A., ... & Zhang, H., “Grains: Generative recursive autoencoders for indoor scenes,” *ACM Transactions on Graphics (TOG)*, 38(2), 1-16, 2019.
- [32] Sun, J. M., Yang, J., Mo, K., Lai, Y. K., Guibas, L., & Gao, L., “Haisor: Human-aware indoor scene optimization via a deep reinforcement learning,” *ACM Transactions on Graphics*, 43(2), 1-17, 2024.
- [33] Zhao, Y., Lin, K., Jia, Z., Gao, Q., Thattai, G., Thomaso, J., & Sukhatme, G. S., “Luminous: Indoor scene generation for embodied ai challenges,” *arXiv preprint* arXiv:2111.05527, 2021.
- [34] Ostosov, A., Wonka, P., & Michels, D. L., “Rlss: A deep reinforcement learning algorithm for sequential scene generation,” *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2219-2228, 2022.
- [35] Ritchie, D., Wang, K., & Lin, Y. A., “Fast and flexible indoor scene synthesis via deep convolutional generative models,” *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6182-6190, 2019.
- [36] Wang, X., Yeshwanth, C., & Nießner, M., “Sceneformer: Indoor scene generation with transformers,” *In 2021 International Conference on 3D Vision (3DV)*, pp. 106-115, IEEE, 2021.
- [37] Sun, Q., Zhou, H., Zhou, W., Li, L., & Li, H., “Forest2seq: Revitalizing order prior for sequential indoor scene synthesis,” *In European Conference on Computer Vision*, pp. 251-268, Springer, Cham, 2025.

〈 저 자 소 개 〉



조 윤 식

- 2021년 한성대학교 컴퓨터공학부 학사
- 2022년 한성대학교 일반대학원 컴퓨터공학과 석사
- 2022년~현재 한성대학교 일반대학원 정보컴퓨터공학과 박사과정
- 관심분야: 가상현실, 증강현실, 컴퓨터그래픽스, HCI
- <https://orcid.org/0000-0003-2118-0904>



이 지 원

- 2018년 부산가톨릭대학교 소프트웨어학과 학사
- 2021년 한국과학기술원 문화기술대학원 석사
- 2024년~현재 한성대학교 일반대학원 정보컴퓨터공학과 박사과정
- 관심분야: 컴퓨터그래픽스, 컴퓨터 비전, VR/AR, LLM
- <https://orcid.org/0000-0001-6848-5110>



조 세 홍

- 1983년 연세대학교 3년 수료
- 1991년 California State Univ. CS 학사
- 1996년 Arizona State Univ. CSE MCS
- 1999년 Arizona State Univ. CSE Ph.D.
- 1999년~2002년 대구대학교 정보통신공학부 조교수
- 2002년~현재 한성대학교 컴퓨터공학부 교수
- 관심분야: 메타버스, 가상현실, 증강현실, 디지털콘텐츠, 멀티미디어, 원격교육, 인공지능, 빅데이터
- <https://orcid.org/0000-0003-3974-8086>



김 진 모

- 2006년 동국대학교 멀티미디어학과 학사
- 2008년 동국대학교 영상대학원 멀티미디어학과 석사
- 2012년 동국대학교 영상대학원 멀티미디어학과 박사
- 2012년~2014년 동국대학교 영상문화콘텐츠연구원 전임연구원
- 2014년~2019년 부산가톨릭대학교 소프트웨어학과 조교수
- 2019년~현재 한성대학교 컴퓨터공학부 부교수
- 관심분야: 컴퓨터그래픽스, VR/AR/MR, 메타버스, 게임 공학 등
- <https://orcid.org/0000-0002-1663-9306>