

AI 기반 이미지 생성 기술의 농업 적용 가능성

윤승리¹ · 이예영¹ · 정은규¹ · 안태인^{2*}

¹서울대학교 농림생물자원학부 대학원생, ²서울대학교 농림생물자원학부 조교수

Agricultural Applicability of AI based Image Generation

Seungri Yoon¹, Yeyeong Lee¹, Eunkyu Jung¹, and Tae In Ahn^{2*}

¹Graduate student, Department of Agriculture, Forestry and Bioresources (Horticultural Science and Biotechnology), Seoul National University, Seoul 08826, Korea

²Assistant Professor, Department of Agriculture, Forestry and Bioresources (Horticultural Science and Biotechnology), Seoul National University, Seoul 08826, Korea

Abstract. Since ChatGPT was released in 2022, the generative artificial intelligence (AI) industry has seen massive growth and is expected to bring significant innovations to cognitive tasks. AI-based image generation, in particular, is leading major changes in the digital world. This study investigates the technical foundations of Midjourney, Stable Diffusion, and Firefly—three notable AI image generation tools—and compares their effectiveness by examining the images they produce. The results show that these AI tools can generate realistic images of tomatoes, strawberries, paprikas, and cucumbers, typical crops grown in greenhouse. Especially, Firefly stood out for its ability to produce very realistic images of greenhouse-grown crops. However, all tools struggled to fully capture the environmental context of greenhouses where these crops grow. The process of refining prompts and using reference images has proven effective in accurately generating images of strawberry fruits and their cultivation systems. In the case of generating cucumber images, the AI tools produced images very close to real ones, with no significant differences found in their evaluation scores. This study demonstrates how AI-based image generation technology can be applied in agriculture, suggesting a bright future for its use in this field.

Additional key words: AI-based image generation, Diffusion model, Generative AI, Image-to-image generation, Prompt engineering, Text-to-image generation

서론

현대 농업은 4차 산업혁명의 급속한 전개와 함께 선진국의 주도 하에 스마트 농업으로 전환되고 있다(Kim 등, 2018). 이는 궁극적으로 센서 기술, 사물인터넷, 데이터 관리 기법, 지능적 의사결정 알고리즘, 로봇 공학 등을 포괄하는 인공지능(AI)의 통합을 목표로 한다(Wakchaure 등, 2023). AI 기술은 식량 안보, 식물 생산 시스템의 최적화, 환경 영향 최소화 영역의 복잡한 문제 해결에 있어 핵심적인 역할을 할 것으로 기대할 수 있다(Alreshidi, 2019; Farooq 등, 2019; Plant 등, 2000).

최근 십년 간 컴퓨팅 하드웨어의 발전, 특히 GPU의 발전에 힘입어 딥러닝을 포함한 AI 기술이 데이터로부터 학습하고 패턴 분류를 수행하는 새로운 패러다임으로 부상하였다(LeCun

등, 2015; Bengio 등, 2021). 또한 PyTorch와 TensorFlow 같은 프레임워크의 지원으로 인공지능과 딥러닝은 특징 추출과 패턴 분류 과정을 원활하게 통합하며, 고품질의 이미지 표현 학습과 시각 인식 작업에서 높은 성능을 달성할 수 있게 되었다(He 등, 2016; Zhai 등, 2022; Lu 등, 2022). 이러한 기술들은 농업 분야에서도 컴퓨터 비전 연구를 수행하는데 있어서 중대한 영향을 주었다(Zhang 등, 2020).

한편, 농업 이미지의 분석과 모델링은 생물학적 다양성과 비구조화된 환경으로 인한 여러 도전을 마주하고 있다(Vougioukas, 2019). 이러한 도전을 극복하기 위해 다양한 영상 조건 하에서 충분한 표본을 포함하는 대규모 데이터셋의 수집이 강조되고 있다(Huang 등, 2023). 그러나 실용적인 애플리케이션에서 필요로 하는 생물학적 재료와 영상 조건의 제약으로 인해, 이러한 데이터셋을 수집하고 주석을 달기까지는 많은 시간과 자원이 소모된다(Lu와 Young, 2020).

생성적 적대 신경망(generative adversarial networks, GANs)

*Corresponding author: tiahn@snu.ac.kr

Received April 16, 2024; Revised April 26, 2024;

Accepted April 29, 2024

과 같은 딥러닝 기반의 이미지 증강 기술은 훈련 데이터의 기본 분포를 학습하여 새로운 데이터를 생성하는 혁신적인 방법을 제공하였다(Creswell 등, 2018; Huang 등, 2018; Goodfellow 등, 2020). 최근 생성형 AI는 대규모 데이터 세트를 기반으로 훈련된 딥러닝 모델을 사용하여 이미지, 텍스트, 코드 등을 생성하거나 변환하는 인공지능 기술로서 괄목할만한 발전을 보여주었다(Dhariwal와 Nichol, 2021). 이러한 기술들은 서로 결합하여 더욱 복잡하고 다양한 데이터 생성 작업을 수행할 수 있는 AI 도구로 개발되고 있으며(Liu 등, 2023), AI 기반 이미지 생성 기술은 현재 디지털 세계의 핵심적인 변화를 주도하고 있다(Goertzel와 Pennachin, 2007; Muller와 Bostrom, 2016; Fjelland, 2020). 이러한 기술은 모델 개발 시 사용되는 기존의 이미지 증강 기술보다 훨씬 더 많은 변화와 정보를 데이터셋에 부여할 수 있는 잠재력을 가지고 있어, 연구 개발에 큰 이점을 제공할 수 있다(Wong 등, 2016; Shorten과 Khoshgoftaar, 2019; Khalifa 등, 2022).

생성형 AI 기술은 농작물 질병 탐지, 잡초 인식, 과일 결합 탐지 등 특정 농업 애플리케이션과 같은 맞춤형 인공지능 시스템의 발전을 도모할 수 있다. 또한 이 기술을 활용하여 대규모 데이터셋의 부족과 같은 기존의 한계를 극복하고, 다양한 영상 변환 기술을 적용함으로써 인공지능 모델의 성능을 향상시킬 수 있다. 본 연구의 목표는 미드저니(Midjourney), 스테이블 디퓨전(Stable diffusion), 파이어플라이(Firefly)와 같은 대표적인 AI 이미지 생성 도구의 원예 작물 및 과일 이미지 생성 능력을 비교하고, 프롬프트를 통한 이미지 개선을 통해 AI 이미지 생성 도구를 효율적으로 농업에 적용하여 실용적인 가이드라인을 제공하는 것이었다.

재료 및 방법

1. 데이터 수집

본 연구에서는 대표적인 시설 재배 작물이면서 각 작물이 가진 독특한 잎과 과실의 형태가 뚜렷이 구분되어 생성된 이미지를 비교하기에 용이한 토마토(*Solanum lycopersicum* L.), 파프리카(*Capsicum annuum* L.), 오이(*Cucumis sativus* L.), 딸기(*Fragaria × ananassa* Duch.)를 대상으로 선정하였다. 생성 이미지와 실제 이미지를 비교하기 위하여, 농촌진흥청에서 공개한 각 작물의 대표 사진을 수집하였다(RDA, 2013).

2. AI 이미지 생성 도구 선정

AI 이미지 생성 도구들은 고유의 시각적 스타일과 처리 기능을 가지고 있어(Derevyanko와 Zalevska, 2023), 동일한 프롬프트에도 불구하고 전혀 다른 결과물을 생성한다(Fig. 1).



Fig. 1. Conceptual representations of future agriculture generated by Midjourney, Stable diffusion, and Firefly, showcasing images integrated with the city (A) and sustainable agricultural practices (B) and the prompts used for generating these concepts (C).

본 연구에서는 기존 연구에서 보고된 생성형 AI의 특성을 바탕으로, 실제 원예 작물과 유사하지만 기존에 존재하지 않는 원예 작물의 이미지를 생성할 수 있는 AI 이미지 생성 도구 사용하였다. 이를 위해 다양한 분야의 많은 이용자를 보유하고 있는 미드저니, 스테이블 디퓨전, 그리고 파이어플라이를 선정하였다. 이 도구들은 각각 그 특성이 뚜렷하여 연구에 적합하다고 평가되었다(Borji, 2022; Jie 등, 2023; Kwon, 2024).

2.1 미드저니(Midjourney)

미드저니(Midjourney Basic Plan, Midjourney, San Francisco, CA, US)는 유료 구독 모델을 통해 서비스되며, 텍스트 프롬프트를 기반으로 이미지를 생성한다(Anyatasia, 2023). 사용자는 자연어로 설명을 입력하고, 미드저니는 이를 바탕으로 시각적 콘텐츠를 생성한다(Reviriego와 Merino-Gómez, 2022). 높은 창의성과 예술적 스타일의 이미지를 생성하는 데 강점을 가지며, 특히 복잡하고 추상적인 아이디어를 시각화하는 데 유용하여, 예술 및 상업 작품 생성에 널리 활용된다(Borji, 2022).

2.2 스테이블 디퓨전(Stable Diffusion)

스테이블 디퓨전(Stability Diffusion, Stability AI, London, LHN, GB)은 커스터마이징 이미지 생성에 오픈소스 프로젝트로 서비스된다. 사용자는 구글 코랩 혹은 로컬 환경에서 실

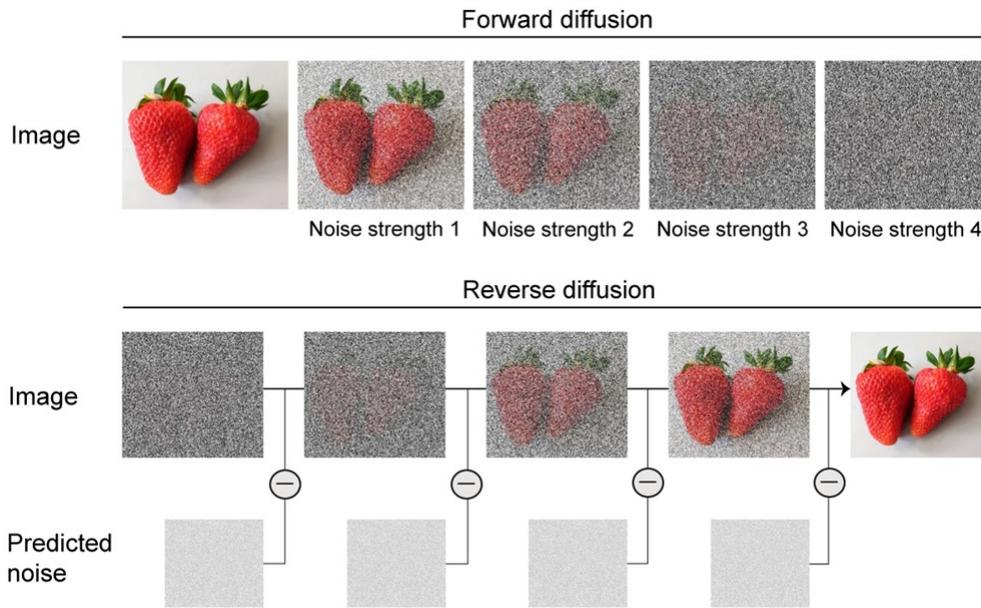


Fig. 2. Image generation process of the diffusion model, illustrating forward diffusion with increasing noise levels and reverse diffusion where noise is reduced based on predictions.

행할 수 있으며, API를 통한 접근에는 제한적으로 비용을 지불해야 한다. 사용자는 다양한 파라미터(예: Seed, Sampling method, CFD scale, Lora, VAE, Checkpoint)를 조정하여 생성 결과의 세부 사항을 제어하고 고해상도의 이미지를 생성할 수 있으며, 다양한 주제에 따른 이미지 생성에 높은 유연성을 가진다. 스테이블 디퓨전의 이미지 생성 원리는 순방향 확산 및 역방향 확산으로 설명할 수 있다. 물 속에서 잉크가 확산(diffusion)되어가듯이 학습 이미지에 잡음(noise)을 조금씩 추가해 무작위 잡음 이미지를 생성한 뒤, 신경망 모델에게 잡음에 따른 이미지들을 학습시키고 잡음을 예측하게 하여, 최종적으로 예측 잡음을 제거해가며 이미지를 생성한다(Fig. 2; Dehouche와 Dehouche, 2023).

2.3 파이어플라이(Firefly)

Adobe 사의 파이어플라이(Firefly, Adobe, San Jose, CA, US)는 크리에이티브 클라우드 서비스의 일부로 제공된다. 어도비의 구독 모델을 통해 접근할 수 있으며, 특정 기능은 무료 사용자에게도 제한적으로 제공된다. 파이어플라이는 사용자가 입력한 텍스트 프롬프트를 기반으로 이미지, 그래픽 디자인 요소, 패턴 등을 생성하며, 디자인 작업을 간소화하고, 사용자가 쉽게 창의적인 콘텐츠를 생성하는 것을 목표로 한다. 포토샵, 일러스트레이션, 애프터 이펙트와 같은 다양한 툴들을 포함하는 통합 어도비 생태계를 통해 접근성이 높으며, 특히 디자인과 시각 예술 분야에서의 적용에 강점을 보인다.

3. 프롬프트 설계 및 이미지 최적화

Liu와 Chilton(2022)는 프롬프트 구성 시 주제(subject)와 스타일(style)을 구체적으로 지정하는 것이 중요한 것으로 보고한 바 있다. 또한, 이미지-이미지 생성 기법은 텍스트-이미지 생성 방법에 비해 이미지 선명도와 품질을 향상시키고 정확하고 현실적인 이미지를 생성할 수 있다(Sapkota 등, 2024). 본 연구에서는 원예 작물과 과일 이미지 생성 작업을 반복적으로 수행하며, 실제 이미지와 유사하도록 개선하는 과정을 거쳤다. 이는 생성 이미지의 질을 높이기 위해 프롬프트의 주제와 스타일에 대한 구체성을 높이는 작업으로, 프롬프트가 중치에 따른 이미지 변화를 세심하게 포착하는 것이 수반된다. 이후 목표로 하는 이미지를 ‘작물명’, ‘흰색 배경’, ‘사진과 같은’, ‘사실적인’, ‘태양광’, 및 ‘온실’이라는 프롬프트 맥락 하에 점진적으로 개선한 뒤, 생성 이미지들을 비교하였다.

4. 모델 성능 평가 및 통계분석

생성 이미지의 품질을 정량적으로 평가하기 위해, OpenAI에서 개발한 CLIP(Contrastive Language-Image Pre-training) 모델을 활용하였다(Radford 등, 2021). 이 모델은 대규모 이미지와 관련된 텍스트 캡션 데이터를 학습하여, 이미지와 텍스트 간의 의미적 관계를 파악하도록 설계되었다. 본 연구에서는 생성된 이미지와 실제 이미지 간의 의미적 유사성을 측정하였다.

이 연구는 미드저니, 스테이블 디퓨전, 그리고 파이어플라

이를 사용하여 딸기 이미지를 세 단계의 프롬프트에 따라 각각 10장씩 생성하였다. 각 단계에서 30장씩 총 90장의 딸기 이미지를 생성하여 비교하였다. 각 이미지는 10장의 실제 딸기 이미지와 비교되었으며, 900개의 CLIP 점수를 계산하여 평가했다.

사전 훈련된 CLIP 모델을 사용하여 각 이미지의 특징 벡터를 추출한 다음, 생성된 이미지와 실제 이미지 사이의 코사인 유사성(cosine similarity)을 계산하였다. 이 유사성은 두 벡터 간의 각도를 기반으로 하여 의미적 가까움을 정량화한다. Python의 `scipy.stats` 및 `statsmodels.stats.multicomp` 라이브러리를 이용한 ANOVA 분석으로 900개의 CLIP 점수의 통계적 유의성을 평가했다.

또한, 딸기 이미지 생성에서 가장 우수한 성능을 보인 세 번째 단계의 프롬프트를 사용하여, 각 모델로부터 30장의 오이 이미지를 생성했다. 이들은 10장의 실제 오이 이미지와 비교되었으며, 이를 통해 총 300개의 CLIP 점수를 계산하여 통계적 유의성을 평가했다.

결과 및 고찰

1. AI 기반 시설원예작물 과실 시각화

시설원에 작물인 토마토, 딸기, 파프리카, 오이의 이미지 생성 결과는 Fig. 3에 나열된 바와 같다. 각 도구들은 실제 과실 이미지와 매우 유사한 이미지를 생성하였다. 미드저니는 특히 줄기와 형태의 구조적 세부 사항, 표면의 결, 요철 및 주름을 세밀하게 재현하며 일관되고 뛰어난 이미지를 제공하였다. 그러나 이러한 도구들은 때때로 현실성이 떨어지는 낮은 품질의 이미지를 생성할 수 있다(Fig. 3E).

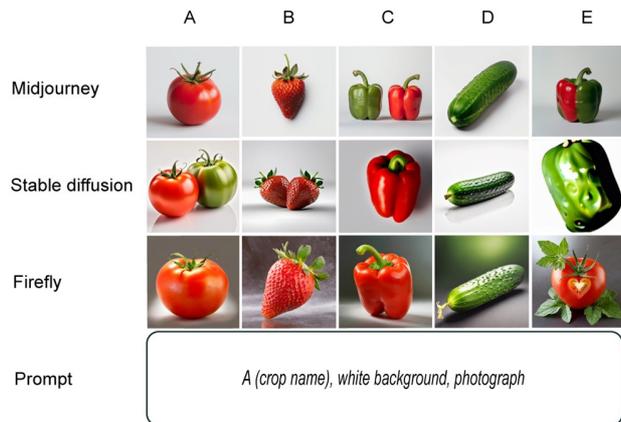


Fig. 3. AI-generated images of fruits: tomato (A), strawberry (B), sweet pepper (C), cucumber (D), and an example of the lowest quality image (E).

이미지 생성 기술은 정밀 농업에서 유전적 개선을 위한 표현형 연구와 같은 현대적 과제를 해결하는 데 중요한 역할을 할 수 있다(Zingaretti 등 2021). AI 도구들은 과실의 외형—특히 모양, 색상, 질감을 비용 효율적이고 신속하게 특성화하는 데 사용될 수 있다. 이는 복잡한 다변량 데이터로서의 외형을 분석하는 자동화된 파이프라인을 제공함으로써, 소비자의 선호도에 기반한 과실의 외적 품질을 효과적으로 평가하는 데 기여할 수 있다(Brewer 등, 2006; Gehan 등, 2017; Feldmann 등, 2020).

2. AI 기반 시설원예작물 시각화

대표적인 시설원에 작물인 토마토, 딸기, 파프리카, 오이의 이미지 생성 결과는 다음과 같다(Fig. 4). 농촌진흥청에서 공개한 실제 작물 이미지들은 각각의 AI 도구가 실제 작물 이미지와 얼마나 유사하게 재현하는지를 평가하는 데 중요한 기준이 된다(RDA, 2013).

미드저니는 빛과 그림자를 매우 사실적으로 처리하여, 자연광 아래에서 자라는 작물의 이미지를 성공적으로 재현해냈으나, 토경 혹은 수경재배, 줄기 유인, 재배 베드 등 작물이 온실

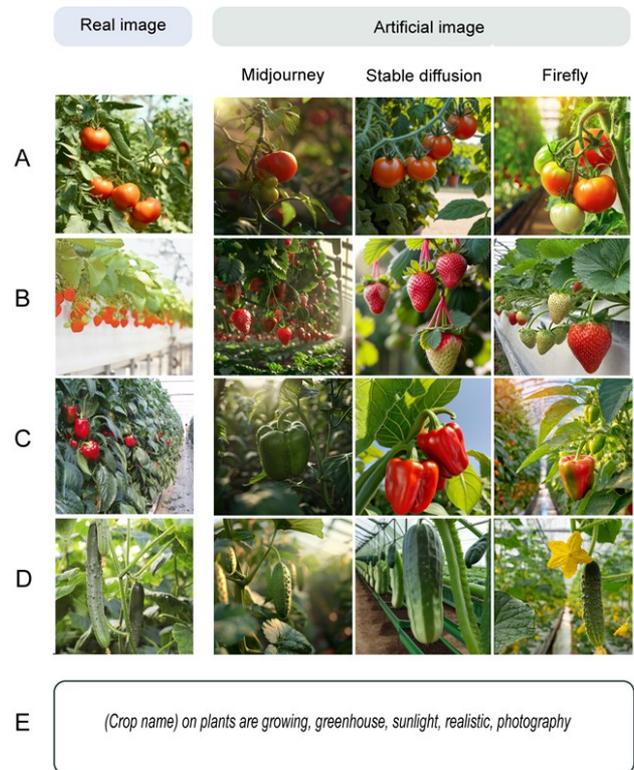


Fig. 4. Comparison of real and AI-generated images of horticultural crops: tomato (A), strawberry (B), sweet pepper (C), cucumber (D), with the prompt used for generation shown in (E).

에서 재배되는 특정한 환경적 맥락을 완전히 반영하지는 못하였다. 스테이블 디퓨전 역시 생성한 색상과 뚜렷한 텍스처로 작물의 형태를 잘 표현했으나, 온실 환경의 묘사는 다소 부족했고 특히 오이 작물을 잘 구현하지 못하였다(Fig. 4D).

반면에 파이어플라이는 실제 온실 재배 작물의 이미지를 가장 사실적으로 묘사하였고, 세밀한 조직감과 색감으로 인해 실제 이미지보다 더 현실적인 느낌을 제공한다. 특히 온실의 재배 베드, 수경재배 시스템의 수직 유인 형태, 토마토, 딸기, 파프리카 과실의 착색 과정(Fig. 4A, B, C), 오이 작물의 화방(Fig. 4D) 등을 구체적으로 재현하였다. 파이어플라이 생성 이미지는 태양광에 따른 작물의 수광태세, 작물의 생육 발달 과정, 표현형 분석, 수확 로봇 디자인 등 식물 과학 연구의 확장 가능성에 대해 시사하는 바가 있다.

3. AI 기반 이미지 생성의 한계점

한편, 이미지 생성 도구들은 사물의 수를 계산하거나, 특정한 물리적 형태나 구조를 정확히 이해하고 재현하는 능력이 제한적이다(Wasielewski, 2023). 이 AI 도구들이 2차원의 이미지 데이터로 훈련되어, 특정한 층수를 가진 건물이나 정확한 수의 손가락을 일관되게 생성하는데 어려움을 겪는다. 또한 복잡한 구조의 3차원 공간, 예를 들어 ‘온실’이나 ‘수경 재배 시스템’을 정확히 생성하지 못하는 경우가 많다(Stöckl, 2023). Stöckl(2023) 연구에 따르면, 스테이블 디퓨전으로 생성된 온실 이미지는 유리나 작물이 포함된 건물로 나타나지만, 온실의 구조적 형태나 식물 생산 시스템의 세부적 묘사는 매우 부족하다.

이에 따라, 실제 농업 환경과 같은 특수한 맥락에서의 이미지 생성은 여전히 한계를 지닌다. 각 도구들에 대한 농업 분야의 훈련 데이터셋이 매우 제한될 뿐 아니라, 그것들이 대표하는 실제 사물의 경험이나 맥락을 이해하지 못하기 때문이다. 농업은 환경과 생물이 상호작용하는 복잡하고 다층적인 시스템이다. 따라서 농업 환경의 의미나 맥락은 AI가 반영하기 어려운 제한 요소로 남아있다. 이러한 한계에도 불구하고 AI 기반 이미지 생성 기술의 발전은 가속화되고 있다. 텍스트-이미지 생성뿐 아니라 이미지-이미지, 이미지-동영상, 텍스트-3D 이미지, 가상공간 3D 객체 생성 기술 등 다양한 영역에서 활발한 연구가 이루어지고 있다(Chang 등, 2014; Kim 등, 2020; Or-EI 등, 2022; Poole 등, 2022). 특히, OpenAI가 개발 중인 대형언어모델 기반 텍스트-동영상 생성기 ‘Sora’는 AI 기반 이미지 생성 기술의 큰 도약을 가져올 것으로 기대되고 있다(Liu 등, 2024).

4. 프롬프트 개선 및 이미지 최적화

본 연구에서는 딸기를 대상으로 프롬프트를 점진적으로 추가하고 개선하여, 딸기의 과실 이미지와 시설 딸기 재배 시스템을 실제와 가깝게 생성하는 과정을 포함하였다(Figs. 5, 6). 딸기 과실 이미지 생성의 경우, 세 단계의 프롬프트 설정을 통해 이미지의 품질을 개선하였다. 각 모델의 이미지 생성 결과, ‘a/an(해당 작물) fruit’, ‘white background’, ‘realistic’, ‘photograph’라는 프롬프트를 사용 시, 실제와 매우 유사한 과실 이미지를 생성되었다.

시설 딸기 재배 시스템의 경우, 초기 단계의 간단한 ‘Strawberry’ 프롬프트로부터 시작하여, 다양한 맥락을 추가하여 상세한 프롬프트 지시를 통해 정교한 이미지를 생성하고자 하였다. 최종적으로 ‘Strawberry plants’, ‘realistic’, ‘sunlight’, ‘greenhouse’이라는 프롬프트에 ‘레퍼런스 이미지’(Fig. 4B; RDA, 2013)를 활용하여, 시설 딸기 재배 시스템의 정확한 구조와 조건을 반영하는 사실감 있는 이미지를 생성할 수 있었다.

프롬프트 엔지니어링은 광범위한 실험과 시행착오를 통해 학습되는 직관적이지 않은 기술로 알려져 있다(Liu와 Chilton, 2022; Oppenlaender 등, 2023). 이 기술을 통해 텍스트 프롬프트에 특정 키워드와 핵심 문구를 추가함으로써 이미지의 미적 품질과 주관적 매력을 향상시킬 수 있음이 밝혀졌다. 이는 이미지 생성 시스템을 특정 방향으로 유도하고 결과 이미지를 조정하는 체계적인 연구로 이어지고 있다(Pavlichenko와 Ustalov, 2023; Oppenlaender, 2023).

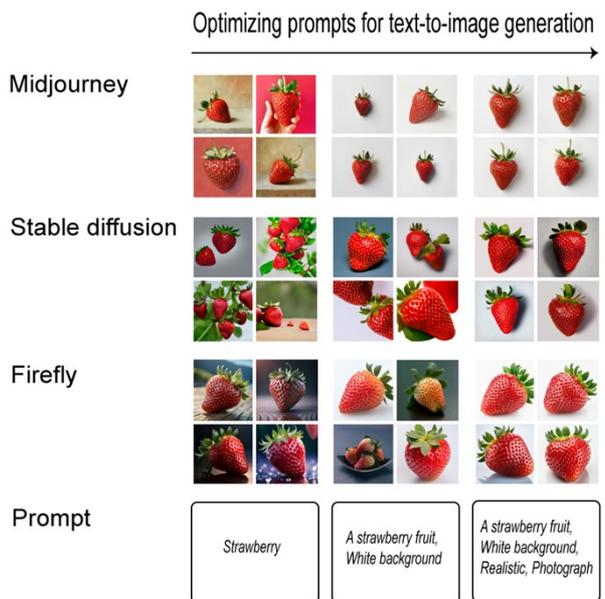


Fig. 5. Progressive improvement of AI-generated strawberry fruit images, optimizing prompts with generative AI tools MidJourney, Stable Diffusion, and Firefly.

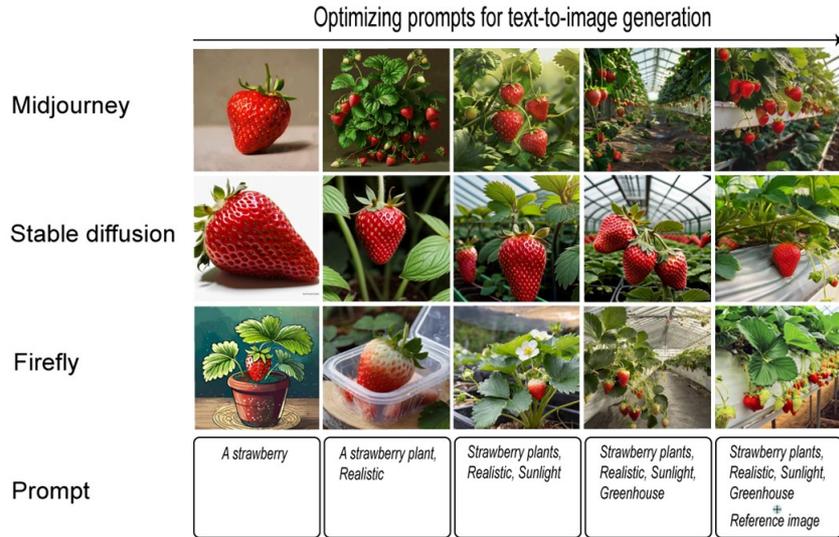


Fig. 6. Progressive improvement of AI-generated strawberry plant images, optimizing prompts and reference images with generative AI tools Midjourney, Stable Diffusion, and Firefly.

5. 모델 성능 비교 및 농업 분야 활용 잠재력

본 연구에서는 AI 생성 도구 미드저니, 스테이블 디퓨전, 그리고 파이어플라이가 생성한 딸기 및 오이 이미지의 품질을 실제 과일 이미지와 비교함으로써 정량적으로 평가하였다 (Figs. 7, 8). OpenAI의 CLIP 모델을 사용하여 각 생성 이미지와 실제 이미지 간의 의미적 유사성을 측정함으로써, 도구 별 성능을 비교할 수 있는 기반을 마련하였다. CLIP 점수는 1에 가까울수록 실제 이미지와 유사성이 높다는 것을 의미한다. 실험 결과, 1단계에서는 스테이블 디퓨전이 0.86, 파이어플라이가 0.85의 평균 CLIP 점수를 기록하였고, 2단계와 3단계로 프롬프트를 개선하면서 모든 모델의 성능이 점진적으로 향상되었다(Fig. 7). 특히 3단계에서는 미드저니와 파이어플라이가 각각 0.93, 0.9로 높은 평균 CLIP 점수를 기록하며 성능이 우수함을 나타냈다. 한편, 스테이블 디퓨전에서는 2단계와 3단계의 CLIP 점수가 유의미한 차이를 보이지 않았는데, 이는 스테이블 디퓨전의 이미지 생성 능력이 프롬프트뿐만 아니라 Check point, Lora, Seed, Sampling method 등의 다양한 파라미터에 크게 의존하기 때문이다.

한편, 오이 이미지 생성에 있어서 총 300개의 CLIP 점수 분석을 통해 얻어진 통계적 유의성 평가에서는 세 도구 간에 유의미한 차이가 발견되지 않았다. 이 결과는 각 생성 도구가 만들어낸 이미지들이 실제 이미지와 유사한 의미적 유사성을 보임을 나타내며, 동시에 각 도구의 고유 스타일과 특성을 반영하고 있음을 보여준다. 전반적으로 이러한 결과는 시설원예 재배 시스템에서의 활용 가능성을 긍정적으로 전망한다.

Borji(2022)는 미드저니, 스테이블 디퓨전, DALL-E 2와 같은 이미지 생성 도구의 얼굴 생성 능력을 정량적으로 비교하였고 스테이블 디퓨전이 사람 얼굴을 잘 생성한다고 평가하였다. Jie 등(2023)은 미드저니와 스테이블 디퓨전의 AI 캐릭터 생성 능력을 비교 분석하였고, 두 도구 모두 독특한 특징을 가진 우수한 캐릭터를 생성했다고 밝혔다. 현재 미드저니, 스테이블 디퓨전, 파이어플라이는 각각 고품질의 이미지 생성 능력으로 인해 다양한 사용자 요구와 선호, 파라미터 개인화, 편의성 등 여러 이유로 널리 사용되고 있다(Kwon, 2024). AI 이미지 생성 기술의 빠른 발전은 디자이너들의 전통적인 생산 한계를 넘어서 무한한 콘텐츠 생성 능력을 가능하게 하고 있으며, 이는 미디어, 교육, 엔터테인먼트, 마케팅 및 과학 연구에 이르는 다양한 분야에서 응용될 수 있는 확장성을 보여준다(Yin 등, 2023).

이전 연구에서는 표준화된 환경에서 과실의 발달 과정, 내·외부의 형태 및 색상을 일관되게 측정하고 분석함으로써, 과실의 중요한 산업적 특성을 정량화할 수 있음을 보여주었다(Zingaretti 등, 2021; Bird 등, 2022). 생성 이미지는 작물의 생육 발달 과정을 한눈에 볼 수 있도록 시각화할 수 있고, 재배 과정을 모니터링 자동화를 위한 학습 이미지를 제공할 수 있다. 과실의 다양한 품질과 관련해서는, 과실의 상품성(Bird 등, 2022)을 대표하는 표준 이미지를 생성하는 용도로 활용될 수도 있다. 농업 교육에서는 AI 기반 생성 이미지가 실제와 거의 흡사한 작물이나 농장 환경의 시각적 자료를 제공하여 학생들이 이론과 실제를 연결하는 데 도움을 줄 수 있으며, 농업

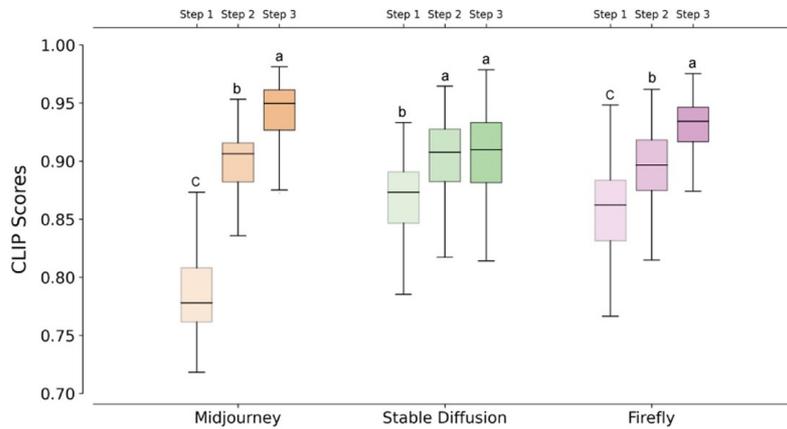


Fig. 7. Quantitative evaluation of CLIP scores showing the effect of prompt handling and added context with generative AI tools Midjourney, Stable Diffusion, and Firefly. Boxplots illustrate distribution and quartile ranges. Alphabetic markers denote significant differences between prompt stages within each model, analyzed using ANOVA and Tukey’s HSD test.

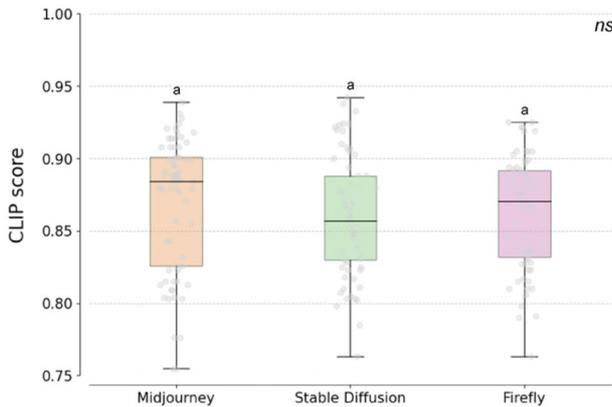


Fig. 8. Quantitative analysis of CLIP scores for Midjourney, Stable Diffusion, and Firefly. Boxplots illustrate distribution and quartile ranges, while individual scatter points illustrate data variability (ns: no significant difference by ANOVA).

관련 교재 개발에도 사용될 수 있어 교육의 질을 더욱 향상시킬 수 있다(Derevyanko와 Zalevska, 2023). 또한 특정 환경 조건이 작물 생장에 미치는 영향을 모델링하고 시뮬레이션하여 환경과 작물의 상호작용을 분석하는데 확장을 기대해볼 수 있다. 더 나아가, 생성형 이미지는 과실의 3D 이미징 변환(Wu 등, 2017)을 통해 과형의 균일성을 평가하고 새로운 유전자형에 따른 표현형을 예측하는 강력한 도구로 활용될 수 있다. 종합적으로, 본 연구에서 생성된 이미지의 시각적 품질과 평가 지표에 대한 고찰은 추후 연구를 통해 농업 및 식물 과학 분야에서 심화 활용 및 범위 확장에 대한 잠재력에 대해 시사하는 바가 있다.

적 요

2022년 ChatGPT 출시 이후, 생성형 AI 산업은 엄청난 규모로 성장하였으며, 인지 작업에 혁신을 가져올 것으로 기대되고 있다. 특히 AI 기반 이미지 생성 기술은 현재 디지털 세계의 핵심적인 변화를 주도하고 있다. 본 연구는 대표적인 AI 이미지 생성 도구인 미드저니, 스테이블 디퓨전, 그리고 파이어플라이의 기술적 원리를 분석하고, 이미지 생성 결과를 비교함으로써 그 유용성을 평가하였다. 실험 결과, 이 AI 도구들은 대표 시설원에 작물인 토마토, 딸기, 파프리카, 오이의 과실 이미지를 실제와 유사하게 재현하였다. 특히 파이어플라이는 실제 온실 재배 작물 이미지를 매우 사실적으로 묘사하는 능력을 보여주었다. 그러나 모든 도구들은 작물이 자라는 온실의 환경적 맥락을 완전히 반영하는 데에 있어서 다소 한계를 보였다. 프롬프트 개선 및 레퍼런스 이미지를 활용하여 딸기 과실 이미지와 시설 딸기재배 시스템을 보다 정교하게 생성하는 과정도 포함되었으며, 이러한 접근은 AI 이미지 생성 기술의 세밀한 조정이 가능함을 보여준다. 오이 과실 이미지 생성 능력을 비교한 결과, AI 생성 도구들은 실제 이미지와 매우 유사한 이미지를 생성해냄으로써 이미지 생성 점수(CLIP score)에 있어서 통계적 차이를 보이지 않았다. 본 연구는 AI 기반 이미지 생성 이미지 기술이 농업 분야에 활용될 수 있는 방안을 모색하며, 생성형 AI의 농업에 대한 적용을 긍정적으로 전망한다.

추가주제어: 생성형 인공지능, 이미지 기반 이미지 생성, 인공지능 기반 이미지 생성, 프롬프트 엔지니어링, 텍스트 기반 이미지 생성, 확산 모델

사 사

본 결과물은 농림축산식품부 및 과학기술정보통신부, 농촌진흥청의 재원으로 농림식품기술기획평가원과 재단 법인 스마트팜연구개발사업단의 스마트팜다부처 패키지혁신기술개발사업의 지원을 받아 연구되었음(423001-02).

Literature Cited

- Alreshidi E. 2019, Smart sustainable agriculture (SSA) solution underpinned by internet of things (IoT) and artificial intelligence (AI). *Int J Adv Comput Sci Appl* 10(5):90-102. doi:10.14569/IJACSA.2019.0100513
- Anyatasia F. 2023, Investigating motivation and usage of text-to-image generative AI for creative practitioner. Available via <https://helda.helsinki.fi/server/api/core/bitstreams/4edf6adb-2d67-4047-bf81-ea09a9b940f1/content>
- Bengio Y., Y. Lecun, and G. Hinton 2021, Deep learning for AI. *Commun ACM* 64(7):58-65.
- Bird J.J., C.M. Barnes, L.J. Manso, A. Ekárt, and D.R. Faria 2022, Fruit quality and defect image classification with conditional GAN data augmentation. *Sci Hortic* 293:110684. doi:10.1016/j.scienta.2021.110684
- Borji A. 2022, Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. arXiv preprint arXiv:2210.00586. doi:10.48550/arXiv.2210.00586
- Brewer M.T., L. Lang, K. Fujimura, N. Dujmovic, S. Gray, and E. van der Knaap 2006, Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiol* 141:15-25. doi:10.1104/pp.106.077867
- Chang A., M. Savva, and C.D. Manning 2014, Learning spatial knowledge for text to 3D scene generation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp 2028-2038.
- Creswell A., T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A.A. Bharath 2018, Generative adversarial networks: An overview. *IEEE Signal Process Mag* 35(1): 53-65.
- Dehouche N., and K. Dehouche 2023, What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. *Heliyon* 9(6):e16757. doi:10.1016/j.heliyon.2023.e16757
- Derevyanko N., and O. Zalevska 2023, Comparative analysis of neural networks Midjourney, Stable Diffusion, and DALL-E and ways of their implementation in the educational process of students of design specialties. *Scientific Bulletin of Mukachevo State University Series "Pedagogy and Psychology"* 9(3):36-44. doi:10.52534/msu-pp3.2023.36
- Dhariwal P., and A. Nichol 2021, Diffusion models beat GANs on image synthesis. *Adv Neural Inf Process Syst* 34:8780-8794. doi:10.48550/arXiv.2105.05233
- Farooq M., A. Rehman, and M. Pisante 2019, Sustainable agriculture and food security. *Innovations in Sustainable Agriculture*, pp 3-24. Springer International Publishing. doi:10.1007/978-3-030-23169-9_1
- Feldmann M.J., M.A. Hardigan, R.A. Famula, C.M. López, A. Tabb, G.S. Cole, and S.J. Knapp 2020, Multi-dimensional machine learning approaches for fruit shape phenotyping in strawberry. *GigaScience* 9:1-17. doi:10.1093/gigascience/giaa030
- Fjelland R. 2020, Why general artificial intelligence will not be realized. *Humanit Soc Sci Commun* 7(1):1-9. doi:10.1057/s41599-020-0494-4
- Gehan M.A., N. Falgren, A. Abbasi, J.C. Berry, S.T. Callen, L. Chavez, A.N. Doust, M.J. Feldman, K.B. Gilbert, J.G. Hodge, and J.S. Hoyer 2017, PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ* 5:e4088. doi: 10.7717/peerj.4088
- Goertzel B., and C. Pennachin (Eds.) 2007, *Artificial General Intelligence*. Springer Berlin Heidelberg. doi:10.1007/978-3-540-68677-4
- Goodfellow I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio 2020, Generative adversarial networks. *Commun ACM* 63(11):139-144.
- He K., X. Zhang, S. Ren, and J. Sun 2016, Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 770-778.
- Huang H., P.S. Yu, and C. Wang 2018, An introduction to image synthesis with generative adversarial nets. arXiv preprint arXiv:1803.04469. doi:10.48550/arXiv.1803.04469
- Huang Z., F. Bianchi, M. Yuksekogonul, T.J. Montine, and J. Zou 2023, A visual-language foundation model for pathology image analysis using medical twitter. *Nat Med* 29(9):2307-2316.
- Jie P., X. Shan, and J. Chung 2023, Comparative analysis of AI painting using [Midjourney] and [Stable Diffusion]-a case study on character drawing. *Int J Adv Culture Technol* 11(2):403-408. doi:10.17703/IJACT.2023.11.2.403
- Khalifa N.E., M. Loey, and S. Mirjalili 2022, A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artif Intell Rev* 55(3):2351-2377.
- Kim D., D. Joo, and J. Kim 2020, Tivgan: Text to image to video generation with step-by-step evolutionary generator. *IEEE Access* 8:153113-153122. doi:10.1109/ACCESS.2020.3017881
- Kim J.G., I.B. Lee, K.S. Yoon, T.H. Ha, R.W. Kim, U.H. Yeo, and S.Y. Lee 2018, A study on the trends of virtual reality application technology for agricultural education. *J Bio-Env Con* 27(2):147-157. doi:10.12791/KSBEC.2018.27.2.147
- Kwon D.H. 2024, Analysis of prompt elements and use cases in image-generating AI: focusing on Midjourney, Stable Diffusion, Firefly, DALL·E. *J Digit Contents Soc* 25(2):341-354. doi:10.9728/dcs.2024.25.2.341
- LeCun Y., Y. Bengio, and G. Hinton 2015, Deep learning. *Nature*

- 521(7553):436-444. doi:10.1038/nature14539
- Liu J., Y. Zhou, Y. Li, Y. Li, S. Hong, Q. Li, X. Liu, M. Lu, and X. Wang 2023, Exploring the integration of digital twin and generative AI in agriculture. 2023 15th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), pp 223-228. doi:10.1109/IHMSC58761.2023.00059
- Liu V., and L.B. Chilton 2022, Design guidelines for prompt engineering text-to-image generative models. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 1-23.
- Liu Y., K. Zhang, Y. Li, Z. Yan, C. Gao, R. Chen, Z. Yuan, Y. Huang, H. Sun, J. Gao, L. He, and L. Sun 2024, Sora: a review on background, technology, limitations, and opportunities of large vision models. arXiv preprint arXiv:2402.17177. doi:10.48550/arXiv.2402.17177
- Lu Y., and S. Young 2020, A survey of public datasets for computer vision tasks in precision agriculture. *Comput Electron Agric* 178:105760.
- Lu Y., D. Chen, E. Olaniyi, and Y. Huang 2022, Generative adversarial networks (GANs) for image augmentation in agriculture: a systematic review. *Comput Electron Agric* 200: 107208. doi:10.1016/j.compag.2022.107208
- Muller V.C., and N. Bostrom 2016, Future progress in artificial intelligence: A survey of expert opinion. *Fundamental issues of artificial intelligence*, pp 555-572. doi:10.1007/978-3-319-26485-1_33
- Oppenlaender J. 2023, A taxonomy of prompt modifiers for text-to-image generation. *Behav Inf Technol* 1-14. doi:10.1080/0144929X.2023.2286532
- Oppenlaender J., R. Linder, and J. Silvennoinen 2023, Prompting AI art: an investigation into the creative skill of prompt engineering. arXiv preprint arXiv:2303.13534. doi:10.48550/arXiv.2303.13534
- Or-El R., X. Luo, M. Shan, E. Shechtman, J.J. Park, and I. Kemelmacher-Shlizerman 2022, Stylesdf: High-resolution 3d-consistent image and geometry generation. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp 13503-13513.
- Pavlichenko N., and D. Ustalov 2023, Best prompts for text-to-image models and how to find them. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp 2067-2071. doi: 10.1145/3539618.3592000
- Plant R., G. Pettygrove, and W. Reinert 2000, Precision agriculture can increase profits and limit environmental impacts. *Calif Agric* 54(4):66-71. doi:10.3733/ca.v054n04p66
- Poole B., A. Jain, J.T. Barron, and B. Mildenhall 2022, Dream-fusion: Text-to-3d using 2d diffusion. arXiv preprint arXiv: 2209.14988. doi:10.48550/arXiv.2209.14988
- Radford A., J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger and I. Sutskever 2021, Learning transferable visual models from natural language supervision. In International conference on machine learning, pp 8748-8763. PMLR. doi:10.48550/arXiv.2103.00020
- Reviriego P., and E. Merino-Gómez 2022, Text to image generation: leaving no language behind. arXiv preprint arXiv: 2208.09333. doi:10.48550/arXiv.2208.09333
- Rural Development Administration (RDA) 2013, Available via https://www.rda.go.kr:2360/ptoPtoFrmPrmnList.do?prgId=pto_farmprmnptoEntry
- Sapkota R., D. Ahmed, and M. Karkee 2024, Creating image datasets in agricultural environments using DALL. E: generative AI-powered large language model. arXiv preprint arXiv:2307.08789. doi:10.48550/arXiv.2307.08789
- Shorten C., and T.M. Khoshgoftaar 2019, A survey on image data augmentation for deep learning. *J Big Data* 6(1):1-48.
- Stöckl A. 2023, Evaluating a synthetic image dataset generated with stable diffusion. In International Congress on Information and Communication Technology, pp 805-818. Singapore: Springer Nature Singapore. doi:10.48550/arXiv.2211.01777
- Vougioukas S.G. 2019, Agricultural robotics. *Annu Rev Control Robot Auton Syst* 2:365-392.
- Wakchaur M., B.K. Patle, and A.K. Mahindrakar 2023, Application of AI techniques and robotics in agriculture: a review. *Artif Intell Life Sci* 3:100057. doi:10.1016/j.aillsci.2023.100057
- Wasielewski A. 2023, Midjourney can't count?: questions of representation and meaning for text-to-image generators. *Interdiscip J Image Sci* 37(1):71-82. Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-510407>
- Wong S.C., A. Gatt, V. Stamatescu, and M.D. McDonnell 2016, Understanding data augmentation for classification: when to warp?. In 2016 International Conference on Digital Image Computing: Techniques and Applications, pp 1-6.
- Wu J., Y. Wang, T. Xue, X. Sun, B. Freeman, and J. Tenenbaum 2017, Marrnet: 3d shape reconstruction via 2.5 d sketches. *Adv Neural Inf Process Syst* 30. doi:10.48550/arXiv.1711.03129
- Yin H., Z. Zhang, and Y. Liu 2023, The exploration of integrating the Midjourney artificial intelligence generated content tool into design systems to direct designers towards future-oriented innovation. *Systems* 11(12):566. doi:10.3390/systems11120566
- Zhai X., A. Kolesnikov, N. Houlsby, and L. Beyer 2022, Scaling vision transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 12104-12113.
- Zhang Q., Y. Liu, C. Gong, Y. Chen, and H. Yu 2020, Applications of deep learning for dense scenes analysis in agriculture: A review. *Sensors* 20(5):1520.
- Zingaretti L.M., A. Monfort, and M. Pérez-Enciso 2021, Automatic fruit morphology phenome and genetic analysis: an application in the octoploid strawberry. *Plant Phenomics* 2021:9812910. doi:10.34133/2021/9812910