O2O 상점의 객체 인식을 위한 생성 AI 기반의 진열대 상품 데이터 증강

시종욱[O], 김성영[*]
[O]국립금오공과대학교 컴퓨터·AI융합공학과,
[*]국립금오공과대학교 컴퓨터공학과
e-mail: jwsi425@kumoh.ac.kr[O], sykim@kumoh.ac.kr[*]

Data Augmentation of Shelf Product for Object Recognition in O2O Stores Based on Generative AI

Jongwook Si[O], Sungyoung Kim[*]
[O]Dept. of Computer·AI Convergence Engineering, Kumoh National Institute of Technology,
[*]Dept. of Computer Engineering, Kumoh National Institute of Technology

● 요 약 ●

본 논문에서는 O2O 상점의 자동화에 필수적인 객체 인식 모델의 성능 향상을 목표로, 생성 AI 기술을 이용한 데이터 증강 방법을 제시한다. 제안하는 방법은 텍스트 프롬프트를 활용하여 진열대 상품 이미지를 포함한 다양한 고품질 이미지를 생성할 수 있음을 보인다. 또한, 실제에 더 가까운 상세한 이미지를 생성하기 위한 최적화된 프롬프트를 제안하고, Stable-Diffusion과 DALL-E2의 생성 결과를 통해 비교 분석한다. 이러한 접근 방법은 객체 인식 모델의 성능 향상에 영향을 미칠 것으로 기대된다.

키워드: 생성 AI(Generative AI), 데이터 증강(Data Augmenation), O2O 상점(O2O Store)

I. Introduction

최근 디지털 기술의 발전과 함께 O2O(Online-to-Offline) 상점의 중요성이 증가하고 있다. 이러한 상점들은 온라인과 오프라인의 경계를 허물고 향상된 쇼핑 경험을 제공한다. 이러한 기술은 물품을 자동으로 인식하고 분류하는 과정이 필요하다. 이 과정에서 객체 인식 기술은 핵심적인 역할을 수행한다.

객체 인식 모델의 성능은 모델의 구조나 방법론이 중요하지만, 학습 데이터의 개수를 무시할 수 없다. 학습 데이터의 질과 양은 모델의 정확도와 신뢰성을 결정짓는 요소로, 실제 환경에서의 성능을 크게 좌우한다. 그러나 고품질의 학습 데이터를 수집하는 것은 시간과 비용이 많이 들고, 특히 다양한 상품을 필요로 하는 O2O 상점에서는 더욱 어렵다.

이러한 문제를 해결하기 위해 본 논문에서는 이미지 생성 AI에 주목하고자 한다. 생성 AI는 인공 지능이 데이터를 자동으로 생성함으로써, 데이터 수집의 어려움을 극복하고, 더 다양하고 풍부한 학습 데이터를 제공할 수 있다. 대표적인 생성 AI로 Stable-Diffusion[1]과 DALL-E2[2]가 있다.

Stable-Diffusion[1]과 DALL-E2[2]는 모두 텍스트 기반의 프롬프트로부터 이미지를 생성할 수 있는 공통점이 있다. Stable-Diffusion[1]은 Nvidia가 개발한 기술로 확률론적 모델링 기술을 이용해 안정적인 이미지를 생성하는 기술로 텍스트 입력을 기반으로 한다. DALL-E2[2]는 OpenAI가 개발한 기술로 텍스트 설명을 입력으로 하여 이미지를 생성한다. Stable-Diffusion[1]과 DALL-E2[2]는 모두 다양하며 고품질의 이미지를 생성할 수 있기 때문에 다양한 분야에서 많이 활용되고 있다 [3-5].

본 논문에서는 O2O 상점에서 사용되는 객체 인식 모델의 향상시키기 위하여 생성 AI 기반의 데이터 증강에 대해 소개한다. 이를 위해, 상점의 물품을 올바르게 인식할 수 있도록 생성 AI를 통해 이미지를 생성하고 활용할 수 있도록 프롬프트를 탐구하고 결과를 비교한다.

II. Image Generation of Shelf Product

객체 인식 모델의 학습 데이터로 사용하기 위해서는 올바른 데이터를 생성해야 한다. 특히, O2O 상점 환경에서는 진열대 전면에 위치한 카메라의 시점에서 촬영된 것과 유사한 이미지가 필요하다. 이에 더해, 실제 상점의 진열대에 있는 다양한 상품과 유사하게 생성된 이미지가 중요하다. 이러한 요구사항을 만족시키는 이미지를 생성하기 위해, 본 논문에서는 이와 관련된 최적의 프롬프트를 개발하고 제안한다.

Stable-Diffusion[1]은 텍스트 프롬프트를 통해 시각적 상황을 구체화할 수 있다. 이를 활용하여 'Shelf', 'Various Products'와 같은 특정 생성 대상을 설정할 수 있다. 그리고 상점 환경을 나타내기 위해 'Market'과 'Convenience Store'와 같은 요소를 추가하고 상점의 전면부를 나타내는 'Front View'를 포함한다. 또한, 실제 상황과 유사한 고품질 데이터 생성을 위해 'High-Quality', 'Realistic', 'Detailed'와 같은 부가적인 기술적 요소들을 프롬프트에 추가한다.

DALL-E2[2]는 프롬프트에 상황을 설명할 수 있는 문장을 입력으로 한다. 제안하는 프롬프트는 위에서 언급한 단어들을 기반으로 작성된 문장이며, “A detailed, realistic image of a shelf in a convenience store, viewed from the front. The shelf is fully stocked with various products commonly found”이다.

그림 1은 제안된 프롬프트를 바탕으로 Stable-Diffusion[1]과 DALL-E2[2] 모델이 생성한 이미지 결과를 보여준다. 두 모델 모두 주어진 시나리오에 맞는 데이터를 성공적으로 생성했으나, 결과물 간에 몇 가지 차이점이 있다. Stable-Diffusion[1]은 진열대에 많은 양의 상품이 배치된 이미지를 생성했지만, 상품들의 종류와 색상이 다소 단조로운 경향을 보이고, 상품명이 흐릿하게 나타나거나 깨지는 경우가 있다. 반면, DALL-E2[2]는 보다 다양한 상품과 세부적인 요소들, 특히 상품명의 표현이 더욱 선명하고 정확하게 생성되었다. 이러한 비교를 통해, DALL-E2[2]의 결과물이 Stable-Diffusion[1]보다 현실적인 상황을 더 잘 반영한다고 할 수 있다.

III. Conclusions

본 논문에서는 O2O 상점 자동화를 위해 객체 인식 모델의 성능을 향상시키는 것을 목표로, 다양한 생성 AI 기술을 활용하여 데이터를 증강하는 방법을 제시하였다. 특히 DALL-E2를 이용해 생성된 데이터는 실제 상점의 진열대에 배치된 상품과 상당히 유사한 것으로 나타났다. 따라서, 이러한 증강 데이터를 사용하여 객체 인식 모델을 학습한다면, 큰 성능 향상이 있을 것으로 기대된다.

ACKNOWLEDGEMENT

본 연구는 2023년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임 [S3344882]

REFERENCES

[1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models", Proc. Of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684-10695, 2023.

[2] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, “Hierarchical text-conditional image generation with clip latents”, arXiv preprint arXiv:2204.06125.

[3] J. Si, and S. Kim, "Fine-tuning using dreambooth: Focusing on realistic virtual interviewer generation based on stable-diffusion", Proc. Of Korean Institute of Information Technology Conference, pp. 145-145, Nov. 2023.

[4] J. Si, and S. Kim, "Diffusion model-based data augmentation for defect detection: Inpainting using prompts", Domestic Conference, Proc. Of Korean Institute of Information Technology Conference, pp. 146-146, Nov. 2023.

[5] Y. Ge, J. Xu, B. N. Zhao, L. Itti, V. Vineet, “Dall-e for detection: Language-driven context image synthesis for object detection”, arXiv preprint arXiv:2206.09592.

Fig. 1. Generated of shelf products with AI (Left: Stable-Diffusion[1], Right: Dall-E2[2])