

국방 객체 데이터 부족 문제 해결을 위한 합성데이터 생성

김병현*, 정남훈, 조규태

LIG넥스원

byunghyun.kim@lignex1.com, namhoon.jung@lignex1.com, kyutae.cho2@lignex1.com

Synthetic Data Generation for Military Objects under Data Scarcity

Byunghyun Kim*, Namhoon Jung, Kyutae Cho

LIG Nex1

요약

인공지능(AI) 모델 성능은 학습 데이터의 양과 질에 크게 의존한다. 하지만 국방 분야는 보안상의 제약뿐만 아니라 실제 데이터를 수집하는 것 자체가 어려운 근본적인 한계를 가지고 있다. 본 논문에서는 국방 분야에서 데이터 확보가 어려운 객체를 활용한 AI 모델 성능 향상을 위해 Stable Diffusion 모델을 기반으로 LoRA(Low-Rank Adaptation) 학습을 통한 합성데이터 생성 방법을 제안하였다. 제안된 방법은 소량의 실제 이미지 데이터와 프롬프트를 입력으로, 특정 국방 객체에 대한 다양하고 사실적인 이미지를 생성한다. 이러한 합성데이터 생성 방법은 국방 분야에서 데이터 부족 문제를 완화하여 국방 AI 연구 발전에 활용될 것으로 기대된다.

1. 서론

인공지능(AI) 기술이 발전함에 따라 이미지 분류와 객체 탐지 분야에서 뛰어난 성과들이 보고되고 있다. 그러나 국방 분야에서는 이미지 분류와 객체 탐지 등 AI를 활용한 연구가 상대적으로 어려움을 겪고 있다. 이는 국방 관련 데이터가 보안상의 제약뿐만 아니라, 실제 데이터를 수집하는 것 자체가 어려운 특수한 환경에 놓여있기 때문이다. 따라서 국방 AI 연구에 필요한 고품질 데이터 확보는 중요한 문제이며, 이를 해결하기 위해 다양한 데이터 생성 연구가 활발히 진행되고 있다. 기존 연구에서는 GAN(Generative Adversarial Network)[1]을 활용하여 합성데이터를 생성하거나 Rotation, Flip, Crop 등의 증강 기법을 통해 데이터를 확보하는 연구가 있었다. 하지만 GAN은 세밀한 부분을 구현하기 어렵고, 데이터 증강 기법은 다양성을 확보하는 데 한계가 있다. 따라서 보다 정교하고 현실감 있는 합성데이터 생성 방법이 요구된다. 합성데이터는 수집 비용이 적고 다양한 조건을 반영할 수 있어, 제한된 환경에서도 AI 모델의 유용한 학습 자원이 될 수 있다. 특히 국방처럼 데이터 접근이 제한적인 분야에서 도메인 특화된 데이터 생성을 통해 충분한 학습 효과를 기대할 수 있다. 최근 주목받고 있는 Text-to-image 모델인 Stable Diffusion[2]은 기존 기법들에 비해 뛰어난 이미지 생성 품질과 유연성을 보여주고 있다. 특히 노이즈 제거 과정을 통해 고해상도 이미지를 생성하며, 프롬프트 조건을 활용하여 구체적인 이미지 생성도 가능하다. 또한 LoRA(Low-Rank Adaptation)[3]와 같은 경량화 학습 기법을 적용하여, 적은 수의 데이터만으로도 특정 객체에 특화된 모델을 효율적으로 학습시킬 수 있다.

본 논문에서는 Stable Diffusion 모델을 기반으로 LoRA 학습을 통해 국방 분야에서 데이터 확보가 어려운 객체 이미지를 생성하는 방법을 제안한다. 소량의 실제 이미지 데이터와 프롬프트를 기반으로 다양한 조건에서 현실감 있는 이미지를 생성하고, 정성적 평가를 통해 검증하였다. 이를 통해 국방 AI 개발에 필요한 데이터 부족 문제를 완화하고, 향후 국방 AI 연구 및 실무 적용에 기여하고자 한다.

2. 국방 객체 합성데이터 생성

본 논문에서는 국방 객체 학습 데이터를 생성하기 위해 Stable Diffusion 기반의 합성데이터 생성 방법을 제안한다. 합성데이터 생성 과정은 태그 생성, 프롬프트 구성, 이미지 생성의 세 단계로 이루어지며, 전체 구조는 그림 1과 같다.

2.1. 태그 생성 (Tag Generation)

태그 생성은 LoRA 학습을 위한 핵심 전처리 과정으로, 실제 이미지에 포함된 객체의 의미와 시각적 특징을 텍스트 형태로 명시화하는 단계다. 태그 생성을 위해 소량의 실제 국방 객체 이미지를 수집한 후, WD14 Tagger[4] 모델을 활용하여 각 이미지의 시각 정보를 반영한 태그를 생성한다. 생성된 태그는 LoRA 학습에 사용되며, 이후 추론 단계에서 입력되는 프롬프트에 해당 태그가 등장했을 때 의미 있는 이미지로 재구성될 수 있도록 하는 역할을 한다.

2.2. 프롬프트 구성 (Prompt Construction)

프롬프트는 Text-to-image 모델의 입력으로 사용되는 문장 형태의 조건 정보로서, 어떤 객체를 어떤 장면에서 어떤 스타일로 생성할지를 결정하는 중요한 입력이다. 특정 객체를 LoRA를 통해 학습시킨 후, 해당 객체가 프롬프트에 명시적으로 등장해야만 생성 이미지 내에서 해당 객체가 등장한다. 따라서 프롬프트에는 정확한 객체 또는 태그명이 반드시 포함되어야 한다. 프롬프트는 “A photo of a single <category_name>”와 같이 Simple prompt[5]로 구성하였다.

2.3. 이미지 생성 (Image Generation)

Stable Diffusion 모델을 개선한 SDXL[6]을 활용하여, 주어진 프롬프트로부터 현실감 있는 이미지를 생성한다. 특히, LoRA 학습은 소량의 학습 데이터만으로도 원하는 객체를 생성할 수 있다. 이 단계는 프롬프트에 따라 다양한 배경, 시점, 조건을 가진 객체 이미지를 자동 생성하여 AI 모델 학습에 활용가능한 대규모 데이터셋을 확보하는 데 목적이 있다.

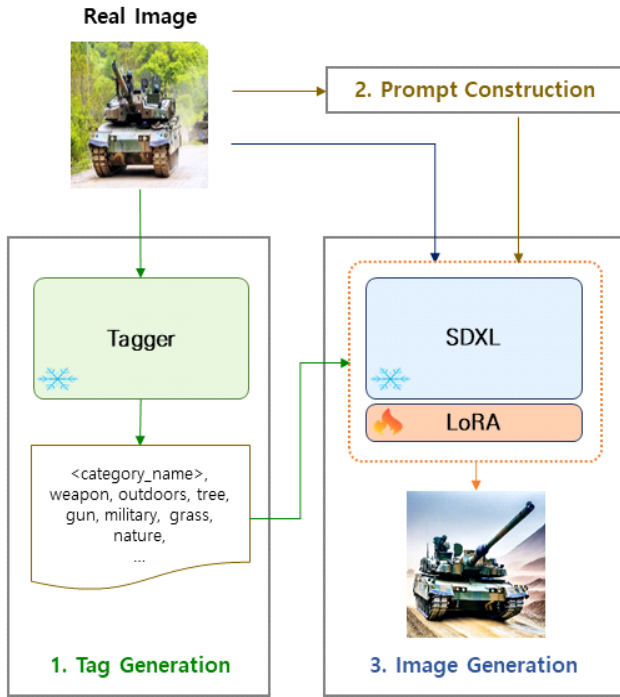


그림 1 국방 객체 합성데이터 생성 방법

3. 실험 결과

실험에서 생성할 국방 객체는 K2 전차로 선정하였고, 데이터가 부족한 상황을 가정하여 인터넷에서 수집 가능한 K2 전차 이미지 10장을 LoRA 학습 데이터로 사용하였다. WD14 Tagger, SDXL 모델은 가중치를 고정했으며, 학습 시 LoRA 가중치만 업데이트하였다. 프롬프트는 Simple prompt를 사용하였고, <category_name>은 “tank”로 설정하였다. 그림 2는 합성데이터 생성의 다양성을 위해 프롬프트에 배경이나 객체의 이동 방향을 포함하여 K2 전차 이미지를 생성한 결과이다. 또한 ControlNet[7]을 적용하여 객체의 위치 및 자세를 제어하고, 학습된 LoRA를 적용한 경우와 그렇지 않은 경우로 나누어 비교하였다. ControlNet 입력은 Visual Prior 중 Depth map을 사용했으며, LoRA를 적용하지 않고 생성한 이미지로부터 추출하였다. 그림 3과 같이, 동일한 Depth map을 사용하여 전차를 생성하더라도 학습된 LoRA를 적용한 경우, 일반적인 전차가 아닌 K2 전차의 특징이 반영된 것을 확인할 수 있다.



그림 2 K2 전차 합성데이터 생성 결과



그림 3 LoRA 적용 유무에 따른 시각적 비교

4. 결론

본 논문에서는 국방 AI 모델 개발에 제약 중 하나인 학습 데이터 부족 문제를 해결하기 위해 Stable Diffusion 기반 모델과 LoRA 학습을 통한 합성데이터 생성 방법을 제안하였다. 제안된 방법은 태그 생성, 프롬프트 구성, 이미지 생성의 세 단계로 구성되며, 소량의 실제 이미지와 프롬프트를 기반으로 특정 국방 객체에 대한 다양하고 사실적인 이미지를 생성한다. 이러한 합성데이터 생성은 기존에 확보가 어려웠던 국방 객체에 대한 학습용 데이터를 대체하거나 보완하는 데 활용될 수 있으며, 결과적으로 이미지 분류 또는 객체 탐지 등 국방 AI 모델의 성능 향상과 연구 개발 비용 절감에 기여할 수 있을 것이다.

참고 문헌

- [1] Goodfellow, Ian J., et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).
- [2] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- [3] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." ICLR 1.2 (2022): 3.
- [4] SmilingWolf, WD 1.4 Tagger, Available: <https://github.com/toriatto/stable-diffusion-webui-wd14-tagger>
- [5] Zhao, Hanqing, et al. "X-paste: Revisiting scalable copy-paste for instance segmentation using clip and stablediffusion." International Conference on Machine Learning. PMLR, 2023.
- [6] Podell, Dustin, et al. "Sdxl: Improving latent diffusion models for high-resolution image synthesis." arXiv preprint arXiv:2307.01952 (2023).
- [7] Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." Proceedings of the IEEE/CVF international conference on computer vision. 2023.