

Stable Diffusion 기반 이미지 생성에 따른 학습 데이터 효율화에 대한 연구

김동현¹, 유대영¹, 정현호¹, 김준화²¹건양대학교 의료인공지능학과²건양대학교 인공지능학과

I. 서론

최근 인공지능 기술의 급속한 발전은 제조업과 의료 분야를 넘어 예술, 교육 등 다양한 영역에서 활발하게 활용되고 있다 [1]. 이러한 발전의 핵심에는 양질의 데이터셋[2]이 있으며, 딥러닝 모델의 성능 향상에는 양질의 데이터셋이 필수적이나, 이를 수집하는 과정은 비용과 시간이 많이 들며 저작권 문제나 조건에 맞는 데이터를 확보하는 데 어려움이 있다 [3].

이러한 문제를 해결하기 위한 대안으로 생성형 AI에 대한 관심이 높아지고 있다. 최근 발전한 딥러닝 기반 이미지 생성 모델, 특히 Stable Diffusion [4]과 같은 기술은 고품질의 이미지를 프롬프트를 통해 인공적으로 생성할 수 있으며, 이를 통해 데이터 수집의 한계를 극복할 수 있는 연구가 진행되고 있다 [5].

본 연구에서는 Stable Diffusion을 활용하여 특정 이미지 데이터셋을 인공적으로 생성하고, 이렇게 생성된 데이터가 실제 데이터셋을 대체하거나 보완할 수 있는지에 대해 탐구하고자 한다. 이를 통해 데이터 수집의 어려움을 완화하고, 다양한 분야에서 학습을 효율화하며 모델의 일반화 성능을 향상시킬 수 있는 방법을 제시하는 것을 목표로 한다.

II. 본론

2.1 Stable Diffusion

Stable Diffusion [4]은 확산 모델(Diffusion Model)을 기반으로 한 이미지 생성 기법 중 하나이다. 기본적으로 노이즈를 추가하고 제거하는 과정을 통해 이미지를 생성하는 방식으로 입력 노이즈에 다양한 조건(텍스트, 이미지 등)을 부여하여 목표로 하는 이미지를 생성할 수 있도록 한다. Stable Diffusion 모델의 핵심은 확률적 과정을 기반으로 이미지를 생성하며 U-Net 구조와 같은 신경망 아키텍처를 통해 기존의 GAN(Generative Adversarial Networks)보다 학습이 안정적이며, 고품질 이미지를 생성할 수 있는 장점이 있다.

2.2 ResNet을 이용한 이미지 분류(Image Classification)

이미지 분류는 주어진 이미지를 사전 정의된 여러 클래스 중 하나로 분류하는 컴퓨터 비전 핵심 문제 중 하나로 다양한 기법들이 발전해 왔으며 최근 딥러닝의 발전과 CNN 등의 기법들이 주요 기술로 자리를 잡았다. 이미지의 특징(Feature)을 추출하고 이를 바탕으로 높은 정확도의 분류 작업을 수행할 수 있는데 그중 ResNet(Residual Networks) [7]은 마이크로소프트에서 개발한 알고리즘으로 [그림 1]과 같이 Skip Connection을 도입함으로써 기울기 소실(Vanishing Gradient) 문제를 효과적으로 해결하였다.

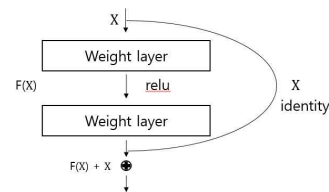


그림 1. 잔차학습(Residual Learning)

III. 실험 결과

3.1 데이터 세트 및 하이퍼 파라미터 설정

본 연구에서는 인공이미지를 실제 데이터 셋으로 사용할 수 있는가를 확인하기 위해, 학습 데이터 셋을 5가지로 분류하였다. 5가지 중 하나는 실제 이미지만 들어가 있는 데이터셋이고 나머지 4가지는 인공 이미지와 실제 이미지의 비율을 다르게 설정하여 만든 데이터셋이다. 인공이미지:실제이미지의 비율은 각각 5:5, 6:4, 7:3, 8:2이다. 사용된 데이터셋은 총 6개의 클래스로 나누어져 있으며, 곰, 개, 고양이, 거북이, 사슴, 독수리로 구성되어 있다. 데이터의 구성은 총 1348개의 학습 데이터, 162개의 검증 데이터, 162개의 테스트 데이터로, 학습 데이터의 약 10%에 해당하는 비율로 검증 데이터와 테스트 데이터를 배분하였다.

인공 이미지는 Stable Diffusion을 사용하여 생성되

었으며, 사용된 프롬프트는 다음과 같다.

프롬프트: “A photorealistic, highly detailed 8K photograph of a turtle in various settings and poses in the sea, shot with a Nikon Z7 II, 70-200mm lens, natural lighting, sharp focus, ultra-high-definition”.

위와 같은 프롬프트로 생성한 인공 이미지와 실제 이미지의 비율을 사용한 데이터셋에서의 모델 성능을 비교한다.

모델 학습에 Optimizer는 Adam, 초기 Learning Rate는 0.0001, Epoch은 25, Batch Size는 32, 손실 함수는 Cross Entropy의 하이퍼 파라미터를 사용하였다.

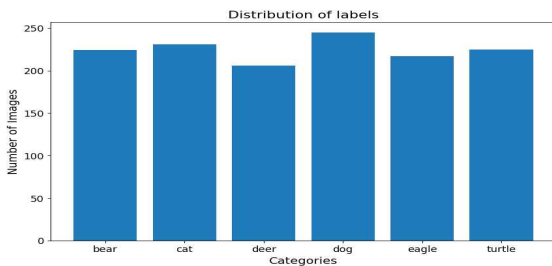


그림2. 사용한 데이터셋의 전반적인 분포



그림 3. Stable Diffusion으로 생성한 인공이미지 예시

3.2 모델 학습

본 연구에서는 ImageNet으로 사전 학습된 ResNet18을 사용하여 비율을 맞춰 나눈 데이터셋을 각각 학습시킨 후 모두 동일한 테스트 데이터를 사용하여 테스트 정확도를 기준으로 성능 비교를 하였다.

[표1]은 위 5가지로 구분한 데이터를 ResNet18 모델을 사용하여 데이터셋을 각각 독립적으로 학습한 결과이다. 원본 데이터셋을 학습한 모델은 98.15%로 가장 높은 정확도를 기록하였고 인공 이미지의 비율이 증가할수록 정확도가 점진적으로 하락하였다. 특히, 가장 많은 인공 이미지의 비율을 사용한 5:5에서 가장 큰 성능 저하가 보였다.

Dataset	Test Accuracy
원본 데이터셋	98.15%
5:5로 나눈 데이터셋	95.06%
6:4로 나눈 데이터셋	97.55%
7:3으로 나눈 데이터셋	94.87%
8:2로 나눈 데이터셋	94.20%

표1. Test Accuracy 데이터 성능 비교

IV. 결론

본 연구에서 Stable Diffusion에서 생성한 인공데이터 이미지를 이용한 ResNet18 기반 Image Classification 모델 연구를 진행하였다. 생성형 AI를 통해 원하는 환경의 인공 데이터를 생성하여 학습을 진행하였을 때 성능의 유지 혹은 상승을 예상하였지만 인공데이터의 비율이 늘어날수록 성능이 저하되었다. 하지만 성능의 저하 폭이 미미한 수준으로 유지되었다. 따라서, 추후 생성형 AI의 이미지 생성 기술이 더 발전하게 된다면 다양한 환경과 원하는 이미지를 생성하여 부족한 데이터셋을 보다 쉽게 수집할 수 있을 것으로 예상된다.

ACKNOWLEDGMENTS

“본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음”(2024-0-00047)

REFERENCES

- [1] Challenges and Opportunities in the Implementation of AI in Manufacturing: A Bibliometric Analysis,“ MDPI, 2023.
- [2] “Artificial Intelligence and Healthcare: A Journey through History, Present Innovations, and Future Possibilities,“ MDPI, 2023..9.
- [3] J.Deng, W.Dong, R.Socher, L.-J.Li, K.Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,“ in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248-255. DOI: 10.1109/CVPR.2009.5206848.
- [4] Alzubaidi, L., et al. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. Journal of Big Data, 10, 46.
- [5] F. A. Croitoru, V. Hondru, R. T. Ionescu and M. Shah, “Diffusion Models in Vision: A Survey,“ in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi:10.1109/TPAMI.2023.3268241
- [6] Wang, Z., Pang, T., Du, C., Lin, M., Liu, W., & Yan, S. (2023). Better Diffusion Models Further Improve Adversarial Training. Proceedings of the 40th International Conference on Machine Learning, PMLR 202:36246-36263..
- [7] He,K.Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. <https://doi.org/10.1109/CVPR.2016.90>