

클래스 불균형 데이터셋의 효과적인 분류를 위한 조건부 확산 모델 기반 데이터 증강 기법

유대현*¹, 김지호*², 이홍철**

Conditional Diffusion Model based Data Augmentation Method for Efficient Classification in Class Imbalance Dataset

Daehyun Yoo*¹, Jiho Kim*², and Hongchul Lee**

본 연구는 4단계 BK21에 의해 지원되었습니다.

요약

클래스 불균형은 분류 모델 성능 저하의 주요한 문제점으로 모델이 소수 클래스에 대한 정보를 충분히 학습하지 못하는데 기인한다. 본 논문은 클래스 불균형 문제를 완화하기 위한 이미지 생성 모델 기반의 두 가지 새로운 방법론을 제안한다. 소수 클래스 이미지를 효과적으로 생성할 수 있는 Diffusion 모델의 클래스와 텍스트 조건부 생성 기법을 활용하였다. 제안한 방법은 기존 데이터 증강 기법 대비 분류 정확도를 평균 1.26, 3.39로 개선하였으며, FID를 21.353, 32.467 만큼 개선하며 기존 데이터 증강 기법보다 효과적인 이미지 생성 성능을 나타내었다. 본 연구는 이미지 분류 및 생성 성능의 향상을 통해 클래스 불균형 문제에 대한 새로운 해결 방법을 제안하며 실제 산업 현장에서의 응용 가능성을 제시하였다.

Abstract

Class imbalance is a major issue in image classification problems as it leads to performance degradation due to the model's inadequate learning of information about minority classes. This paper proposes two methodologies based on Diffusion models to mitigate the class imbalance issue by effectively generating images of minority classes. These methodologies utilize class and text-conditional generation techniques of Diffusion models. This methodology exhibited enhanced classification performance compared to existing imbalanced datasets, improving the accuracy by an average of 1.26 and 3.39 compared to existing data augmentation techniques while also enhancing FID by 21.353 and 32.467, respectively. This paper introduces a novel solution to the class imbalance problem through improvements in image generation and classification performance, suggesting its potential applicability in real-world industrial scenarios.

Keywords

class imbalance, image classification, diffusion model, data augmentation, image generation

* 고려대학교 산업경영공학부
- ORCID¹: <https://orcid.org/0009-0002-3362-8438>
- ORCID²: <https://orcid.org/0000-0003-3733-8702>
** 고려대학교 산업경영공학부 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-4407-0348>

• Received: Feb. 20, 2024, Revised: Apr. 09, 2024, Accepted: Apr. 12, 2024
• Corresponding Author: Hongchul Lee
Dept. of Industrial and Management Engineering, Korea University, Korea
Tel.: +82-2-3290-3767, Email: hlee@korea.ac.kr

1. 서 론

이미지 분류는 다양한 산업에서 다루는 핵심과제 중 하나이다. 대용량의 이미지 데이터가 끊임없이 생성되는 현대 산업에서, 수많은 이미지를 분류하는 것은 정보를 효율적으로 검색하고 활용하게 하는데 필수적이다. 머신러닝 및 딥 러닝 기술의 발전은 이미지 분류 모델의 상당한 발전을 이루어 냈다. 예를 들어, 딥 러닝 모델을 통해 X-ray와 MRI 이미지를 분석하여 병변을 찾아내거나, 유방암 및 기타 종양을 빠르고 정확하게 진단할 수 있다[1][2]. 또한, 이미지 분류 모델은 제조 분야에서도 제품 불량을 판별하거나, 생산 공정에서의 이상 탐지에 사용되고 있다[3][4]. 이 외에도 해충, 패션 이미지 분류, 상품 카테고리 분류 등과 같이 다양한 분야에서 활용되고 있다[5]-[7].

그러나 이미지 분류 작업에서는 데이터셋 구성에 주의가 요구된다. 모델의 학습 과정에서 특정 클래스에 치우쳐 학습되어 학습될 수 있기 때문이다. 이는 분류 문제에서 흔히 발생하는 클래스 불균형 문제로 데이터의 구성이 몇 개의 클래스로 치우쳐 있을 때 오분류의 가능성을 높이고 모델의 편향을 증가시킨다. 이는 다양한 산업에서 활용되는 머신러닝 모델의 성능과 신뢰도를 저하시키는 주요 원인이 된다. 따라서 클래스 불균형을 해결하는 것은 모델의 예측 성능 개선과 모델의 신뢰성 및 신용성을 확보하는데 중요한 요구사항이다[8].

클래스 불균형 문제를 해결하기 위한 다양한 방법들이 제안되어왔다. 대표적인 방법으로 오버샘플링, 다운샘플링 등이 있다. 그러나 이러한 방법들은 기존 데이터 손실이나 과적합 문제를 야기할 수 있으므로 더 효과적인 클래스 불균형 처리를 위한 연구가 지속되고 있다. 최근에는 딥 러닝 기반의 VAE(Variational AutoEncoder)[9]와 GAN(Generative Adversarial Network)[10]과 같은 생성 모델을 활용한 접근법들이 주목받고 있다. 생성 모델의 발전에 따라 다양한 분야에서 데이터 불균형 문제를 해결하기 위해 사용되고 있다[11]-[14]. 그러나 VAE는 생성된 이미지가 흐릿하고 학습이 불안정하며 GAN은 데이터의 다양성을 제대로 반영하지 못하는 한계점

이 있다.

한편, 최근에는 고품질의 이미지를 안정적으로 만들어 내며 VAE, GAN의 한계점을 극복한 Diffusion 모델이 등장하였다. Diffusion 모델은 데이터를 생성하는 모델 중 하나로, 원본 데이터의 분포를 학습하여 새로운 데이터 포인트를 생성하는 데 사용된다[15]. 특히, Diffusion 모델은 복잡하고 다양한 데이터 분포를 정확하게 표현할 수 있어 다양한 응용 분야에서 주목받고 있다.

본 연구에서는 Diffusion 모델의 조건부 생성을 활용하여 클래스 불균형을 해결하는 두 가지 방법을 제안한다. 조건부 생성 기법은 텍스트, 클래스와 같은 추가 정보에 따라 데이터를 생성하는 작업을 의미한다. 첫 번째 제안 방법은 Classifier-free guidance[16]에 클래스 불균형 데이터셋을 학습시켜 데이터의 개수가 부족한 클래스의 데이터를 생성하였다. 이미지를 생성하는 과정에서 생성하고자 하는 클래스 정보를 입력하여 생성한다. 두 번째 제안 방법은 사전 학습된 Stable Diffusion[17]의 텍스트 조건부 생성을 활용해 원하는 클래스의 데이터를 생성하였다. 생성하는 과정에서 텍스트 정보를 입력하여 생성한다. 데이터가 부족한 클래스의 이미지를 생성하기 위해 텍스트 정보는 해당 클래스의 이름(예: cat)을 사용하였다. 본 논문에서는 생성된 데이터와 기존의 클래스 불균형 데이터를 결합하여 균형 잡힌 데이터셋으로 재구성하였다. 이를 통해 이미지 분류 작업을 수행하였으며 사용한 모델은 ResNet18과 ResNet50[18] 이다.

실험은 제안된 두 가지 기법을 적용하여 생성된 이미지를 기존 이미지 데이터 증강연구에서 많이 활용되었던 방법인 Mixup[19]와 CutMix[20]를 비교 방법론으로 활용하였다. 기존 CIFAR10 데이터셋과 생성된 이미지의 유사성을 평가하기 위해 FID(Fréchet Inception Distance) 점수를 사용하여 비교하였으며, 각각 생성된 이미지로 균형 잡힌 데이터셋을 분류 모델에 학습시켜 Accuracy와 F1-score를 비교하였다. 실험 결과, 본 논문에서 제안된 방법이 기존의 증강기법에 비해 FID 성능을 향상하고 CIFAR10에 근사한 이미지를 생성하였으며 분류 성능 또한 유의미하게 향상되었다.

논문의 구성은 다음과 같다. 2장에서는 클래스 불균형 관련 연구들과 논문에서 활용한 Diffusion 모델을 서술한다. 3장과 4장에서는 제안한 방법론과 실험 설계 및 결과를 서술하며, 마지막으로 5장에서는 결론으로 연구의 요약과 추후 연구에 대해 제시하였다.

II. 관련 연구

2.1 Class imbalance

클래스 불균형은 분류 모델 학습 시에 데이터 클래스 간의 빈도가 균일하지 않을 때 발생하는 현상을 의미한다. 모델 학습 과정에서 이러한 불균형은 다수 클래스에 대한 학습은 개선되지만, 소수 클래스에 대해서는 학습이 제대로 이루어지지 않아 모델 성능을 저하할 수 있다. 이를 해결하기 위해 데이터 증강기법이 활용되며, 이는 원본 훈련 데이터셋의 변형을 통해 새로운 데이터를 생성하는 방식이다. 해당 방식은 데이터의 다양성을 증가시키고 과적합을 방지하는 데 효과적이다.

대표적인 데이터 증강기법으로는 회전, 확대/축소, 좌우 반전 등이 있으며, 이를 통해 데이터의 다양성을 증가시키며, 소수 클래스의 데이터를 증강하여 클래스 간 불균형을 해소한다. 또한, 이미지 분류 성능 향상을 위해 설계된 Mixup[19] 기법은 학습 데이터에서 두 개의 샘플 데이터를 혼합(Mix)하여 새로운 학습 데이터를 만드는 기술이다. 두 개의 이미지 쌍 x_i, x_j 와 그에 상응하는 라벨 쌍 y_i, y_j 을 선형적으로 하이퍼 파라미터(λ)에 따라 식 (1)과 같이 혼합한다. 본 논문에서는 $\lambda = 0.6$ 로 설정하였다.

$$\begin{aligned} x_{mix} &= \lambda x_i + (1 - \lambda) x_j \\ y_{mix} &= \lambda y_i + (1 - \lambda) y_j \end{aligned} \quad (1)$$

CutMix[20]는 두 이미지를 겹쳐서 잘라내고 서로 교차시켜 새로운 이미지를 생성하는 방식이다. CutMix는 이미지에 CutMix를 적용할 확률을 결정하는 하이퍼 파라미터(λ)를 사용하며, 두 이미지를 자를 영역을 결정하는 랜덤 이진 마스크(Mask)를 사용하며 식 (2)와 같이 생성된다.

$$\begin{aligned} x_{cutmix} &= Mask \cdot x_i + (1 - Mask) \cdot x_j \\ y_{cutmix} &= \lambda y_i + (1 - \lambda) y_j \end{aligned} \quad (2)$$

최근 연구에서는 VAE[9], GAN[10] 같은 생성 모델을 이용한 데이터 증강기법도 다양하게 연구되고 있는데, 이러한 기법은 데이터의 다양성과 품질을 높이며 데이터 증강기법으로 활용되었다 [21][22]. 더불어, 최근에는 Diffusion 모델이 생성 모델 분야에서 주목받고 있다. 이 모델은 기존의 VAE, GAN보다 복잡한 구조의 이미지를 더 우수한 품질로 생성할 수 있어, 생성 모델의 새로운 가능성을 보여주고 있다.

본 논문에서는 기존에 사용되지 않았던 Diffusion 모델을 활용한 데이터 증강을 통해 클래스 불균형 문제를 해결하고자 한다. 이를 통해, 기존 데이터셋을 개선함으로써 모델의 성능을 향상하는데 기여하였다.

2.2 Diffusion model

Diffusion 모델은 이미지 생성에 사용되는 확률적 방법론 중 하나이다. Diffusion 모델은 원본 이미지에 점진적으로 노이즈를 추가하는 forward process와, 노이즈가 추가된 이미지를 원본 이미지로 복원하는 reverse process로 이루어져 있다. Reverse process를 통해서 forward process에 점진적으로 추가된 노이즈를 제거하여 원본 이미지를 추정하며 학습한다. 그림 1은 Diffusion 모델의 forward process와 reverse process를 표현한 것이다.

이후, Diffusion 모델의 기본 구조를 발전시킨, DDPM(Denoising Diffusion Probabilistic Model)[23]이라는 새로운 모델이 제안되었다. DDPM은 Diffusion 모델의 수식과 모델 학습 절차를 개선하여 높은 품질의 이미지를 생성하였다. DDPM의 forward process는 주어진 $q(x)$ 의 분포를 따르는 원본 이미지 $x_0 \sim q(x)$ 에 점진적으로 노이즈를 더해가는 과정이며, 이는 $t-1$ 시점에서 Markov Chain 방식에 따라 β_t 의 분산을 가진 가우시안 노이즈가 추가되는 방식으로 진행된다. Forward process는 식 (3)과 같이 나타낼 수 있다.

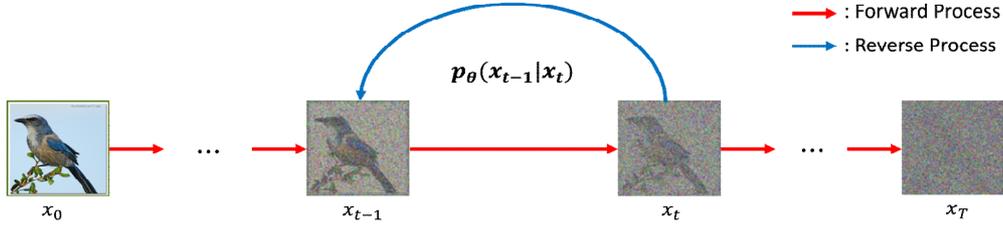


그림 1. Diffusion 과정
Fig. 1. Diffusion process

$$q(x_{1:T}|x_0) := \prod_{t=1}^T q(x_t|x_{t-1}) \quad (3)$$

$$q(x_t|x_{t-1}) := N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

Reverse process에서는 모델은 Denoising을 학습하여, 특정 시점 t 의 이미지 x_t 로부터 x_{t-1} 의 평균 μ_θ 와 분산 Σ_θ 을 예측하며 학습한다. Reverse process는 아래 식 (4)로 표현된다.

$$p(x_{0:T}) = p_\theta(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t) \quad (4)$$

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

Diffusion 모델의 목적은 원본 데이터의 분포 $p_\theta(x_0)$ 를 학습하는 것이며, 이는 식 (5)의 loss를 최소화하면서 진행된다. 식 (5)의 loss를 최소화하는 reverse process를 학습하여 노이즈로부터 데이터를 만들어내는 생성 모델을 만들어낸다.

$$E[-\log p_\theta(x_0)] \leq E_q \left[-\log \frac{p_\theta(x_{0:T})}{q(x_{1:T}|x_0)} \right]$$

$$= E_q \left[-\log p(x_T) - \sum_{t \geq 1} \log \frac{p_\theta(x_{t-1}|x_t)}{q(x_t|x_{t-1})} \right] \quad (5)$$

$$=: L$$

DDPM은 기존의 loss 함수가 복잡한 데이터 분포와 관련된 문제를 효과적으로 처리하기 어렵다는 한계점을 해결하기 위해 ELBO(Evidence Lower Bound)를 도입하였다. ELBO의 KL-Divergence를 사용하여 분포의 차이를 계산하고, 이를 통해 모델의 loss 함수를 최적화하였다. 아래 식 (6)는 ELBO를 적용한 loss 함수이다.

$$\log p(x) \geq$$

$$E_q(x_1|x_0) [\log p_\theta(x_0|x_1)$$

$$- D_{KL}(q(x_T|x_0)|p(x_T)) \quad (6)$$

$$- \sum_{t=2}^T E_{q(x_t|x_0)} [D_{KL}(q(x_{t-1}|x_t, x_0)|p_\theta)]]$$

$$= L_0 - L_T - \sum_{t=2}^T L_{t-1}$$

또한, DDPM에서는 inductive bias를 늘려 모델의 학습과 최적화 과정을 개선하였다. 최종적인 DDPM 식 (7)로 표현할 수 있으며 은 추정되는 노이즈를 의미한다.

$$L_t^{simple} =$$

$$E_{x_0, t, \epsilon} \left[\left\| \epsilon - \epsilon_\theta(\sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\epsilon, t) \right\|^2 \right] \quad (7)$$

따라서, 본 논문에서 Diffusion model을 활용한 이미지 생성을 통해 클래스 불균형을 해결하고자 한다.

III. 방법론

본 논문에서는 클래스 불균형 데이터셋을 균형 잡힌 데이터셋으로 재구성하기 위해 이미지 생성 방법론 중 하나인 Diffusion 모델을 사용하였다. 본 논문에서는 Diffusion 모델의 조건부 생성 기법을 활용하여 소수 클래스에 대한 클래스 불균형의 문제를 해결하고자 한다. 이미지를 생성하여, 클래스 불균형 데이터셋을 균형 잡힌 데이터셋으로 재구성하였다. 첫 번째로 제안한 방법은 Classifier-free guidance[16] 모델을 사용하여 클래스 조건부 생성을 통해 새로운 이미지를 만들었으며, 두 번째 제안 방

법에서는 LDM[17]을 활용해 텍스트 조건부 생성을 통해 이미지를 만들었다. 이를 통해, 데이터가 부족한 클래스의 이미지를 기존의 균형 잡힌 원본 CIFAR10 이미지와 유사하게 재구성할 수 있었다. 이후, 재구성된 데이터셋을 통해 이미지 분류 성능을 확인하였다.

3.1 Classifier-free guidance

DDPM[23]은 뛰어난 생성 능력을 통해 기존의 VAE[9], GAN[10]보다 다양성, 품질 측면에서 우수한 성능을 보여주었다. Diffusion 모델은 UNet[24] 기반의 구조로 구현되었다. 이 구조는 이미지를 생성하는 과정에서 중요한 역할을 하는데, 각 계층에서의 특징 추출을 통해 세밀하고 정확한 이미지 생성이 가능하게 만든다. DDPM 이후에는 Classifier를 추가로 학습해 원하는 클래스의 이미지를 생성할 수 있는 Classifier-Guidance(ADM)[25]와 Classifier를 따로 학습하지 않고, Diffusion 모델에서 학습하는 CFG가 등장하였다. 먼저, ADM은 생성된 이미지의 품질을 향상하기 위해 별도의 Classifier를 학습하는 방법을 제안하였다. 기존 Diffusion 모델의 학습 방법은 동일하나, Classifier로부터 나오는 gradient를 예측한 노이즈에 추가한다. Classifier의 기울기를 조절함으로써 해당 클래스의 이미지를 생성할 수 있었다. 그러나 이러한 모델 구조는 추가적인 Classifier를 학습시켜야 하므로 Diffusion 모델의 학습이 복잡해지고, Classifier는 노이즈가 있는 데이터에 대해 학습해야 하므로 사전학습된 Classifier를 활용할 수 없다는 한계점이 존재했다.

이러한 문제점을 해결한 방법이 CFG이다. CFG는

Classifier의 기울기를 이용한 샘플링 대신, 조건부 Diffusion 모델의 출력값과 비조건부 Diffusion 모델의 출력값 차이의 방향으로 샘플링 하도록 유도한다. 아래 식 (8)은 CFG의 샘플링 함수를 나타낸다.

$$\tilde{\epsilon}_\theta(z_\lambda, c) = (1 + w)\epsilon_\theta(z_\lambda, c) - w\epsilon_\theta(z_\lambda) \quad (8)$$

w 는 guidance scale로 w 가 커질수록 특정 클래스의 이미지만 생성되므로 CFG를 활용하여 원하는 클래스의 데이터를 생성할 수 있다. 본 논문에서는 클래스가 불균형한 CIFAR10 데이터셋을 CFG 모델에 학습시켜 소수 클래스의 데이터를 생성하였다. 그림 2는 제안된 방법론의 전체 과정을 나타낸다. 불균형한 CIFAR10 데이터셋을 CFG 모델에 학습시켜 이미지 데이터를 생성하였다. 이후, 생성된 데이터는 원래의 불균형 데이터셋에 추가하여 최종적으로 균형 잡힌 클래스를 가진 데이터셋으로 재구성하였다.

3.2 Stable Diffusion(LDM)

클래스 정보뿐만 아니라 텍스트 정보를 활용한 GLIDE[26], Imagen[27], DALL-E2[28] 같은 방법론들도 연구되고 있다. 그러나 [26]-[28]의 Diffusion 모델은 픽셀값을 직접 예측하는 Diffusion model의 특성상 고해상도 이미지를 생성하기 위해서는 많은 컴퓨팅 자원을 필요로 하는 한계가 있다. 고해상도 이미지는 이미지의 세부 정보와 패턴을 더 잘 보존하여 정교한 결과를 얻을 수 있다. 따라서 고해상도 이미지를 용언기 위해서는 별도로 Super-Resolution 모델을 학습하여 적용한다[29].

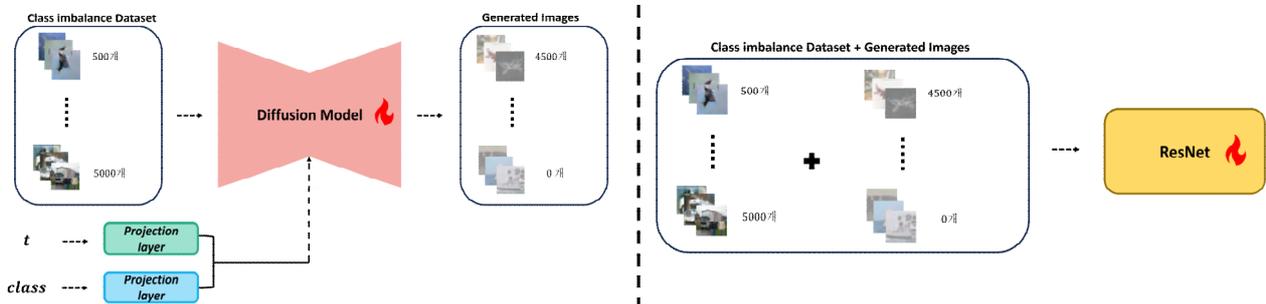


그림 2. Classifier-free guidance 구조를 활용한 이미지 분류 프레임워크
 Fig. 2. Image classification framework utilizing classifier-free guidance

고해상도 이미지 생성은 시각적 세부 사항을 더 정확하게 보존함으로써 더욱 풍부하고 실제적인 결과를 만들어낸다. 이에 따라 데이터 증강과 함께 고해상도 이미지 생성 기술은 다양한 응용 분야에서 효과적인 결과를 나타냈다[30].

Stable Diffusion(LDM)[17]은 이러한 기존 Diffusion 모델의 제약을 극복하기 위해 잠재공간을 활용해 고해상도의 이미지를 생성하고, 더 적은 컴퓨팅 자원으로 효율적인 Diffusion 모델을 개발하였다. 이를 통해, 기존의 잠재공간을 사용하지 않는 GLIDE, Imagen, DALL-E2 보다 적은 컴퓨팅 자원과 시간을 소모하여 뛰어난 품질의 이미지를 생성할 수 있게 되었다.

식 (9)은 LDM의 loss 함수를 나타내며, 이를 통해 픽셀값이 아닌 잠재공간 임베딩 z 가 입력으로 사용된다는 것을 확인할 수 있다.

$$\begin{aligned} L_{DM} &= E_{x, \epsilon \sim N(0,1), t} [\|\epsilon - \epsilon_{\theta}(x_t, t)\|_2^2] \\ L_{LDM} &= E_{\epsilon(x), \epsilon \sim N(0,1), t} [\|\epsilon - \epsilon_{\theta}(z_t, t)\|_2^2] \end{aligned} \quad (9)$$

본 연구에서는 사전 학습된 LDM[17]을 사용하여 클래스 불균형 데이터셋에서 부족한 클래스의 이미지를 생성하였다.

고해상도 이미지를 생성하여 원본 CIFAR10에 근사한 이미지를 생성하고자 하였다. LDM의 생성 과정에 사용되는 텍스트 정보는 클래스 불균형 데이터셋의 각 클래스 이름으로 설정하였다. 그림 3은 LDM을 활용한 이미지 생성 과정을 나타낸 것이다. LDM의 출력 이미지는 512x512 이미지이기 때문에 클래스 불균형 데이터셋의 이미지 사이즈로 축소하여 재구성하였다.

재구성된 이미지는 CFG 방법론과 동일한 방식으로 기존의 클래스 불균형 데이터셋과 결합하여 균형 잡힌 데이터셋을 구성하였다.



그림 3. 텍스트 조건부 생성을 활용한 이미지 생성 프레임워크

Fig. 3. Framework of proposed method with generation

IV. 실험 및 결과

4.1 Dataset

본 논문에서는 CIFAR10 데이터셋을 사용하였다. CIFAR10 데이터셋은 10개의 클래스로 구성되어 있으며, 각 클래스별로 32x32 해상도의 5,000개의 학습 데이터와 1,000개의 평가 데이터로 구성되어 총 60,000개의 데이터셋으로 이루어져 있다.

본 논문에서는 CIFAR10 데이터셋을 클래스 불균형에 대한 실험에 적용하기 위해 클래스별로 데이터 개수가 같은 CIFAR10 데이터를 재구성하여 사용하였다[30]. 클래스 불균형한 CIFAR10 데이터셋을 구성하기 위해 식 (10)을 통해 적용하였다.

$$N_i^{train} = 500 \cdot (i + 1) \quad (10)$$

i 는 각 클래스의 인덱스 번호이며, 클래스별로 점진적으로 개수가 늘어나도록 설정하였다. 즉, 0번 클래스의 데이터 개수는 500개, 9번 클래스의 데이터 개수는 5000개이다. 최종적으로 학습에 사용한 각 클래스별 데이터의 개수는 아래 표 1과 같다. 테스트 데이터셋의 경우 기존 테스트셋을 사용하여 평가하였다.

표 1. 재구성된 CIFAR10 데이터 세트
Table 1. Reconstructed CIFAR10 data sets

Index(i)	Class name	Data number
0	Airplane	500
1	Automobile	1000
⋮	⋮	⋮
8	Ship	4500
9	Truck	5000

4.2 실험 성능 평가 지표

4.2.1 이미지 생성 성능 평가 지표

제안하는 방법론을 통해 생성된 이미지를 통해 재구성된 데이터셋이 기존 CIFAR10 데이터셋과 얼마나 유사한지 확인하기 위해 이미지 생성 성능 평가 지표인 FID(Fréchet Inception Distance)[29]를 사용하였다. FID는 두 이미지 집합 사이의 분포를 비교하고 Fréchet 거리를 계산하여 이미지의 다양성과 충실도를 측정하는 데 활용된다. 이 평가 지표는 두 이미지 집합 사이의 유사도를 정량적으로 보여준다. 즉, 두 데이터셋의 분포가 유사할수록 FID 점수는 낮아지며 기존 CIFAR10 데이터셋과 유사하게 재구성되었다고 볼 수 있다. FID의 계산 방식은 아래 식 (11)과 같다.

$$FID = \frac{\|\mu_X - \mu_Y\|^2}{-Tr(\Sigma_X + \Sigma_Y - 2\sqrt{\Sigma_X \Sigma_Y})} \quad (11)$$

μ_X 와 μ_Y 는 각 데이터셋의 특징 벡터들의 평균을 의미하며, Σ_X 와 Σ_Y 는 각 데이터셋의 특성 벡터의 공분산 행렬을 나타낸다. Tr 은 행렬의 대각 성분의 합을 나타낸다. 본 연구에서 X 는 기존 CIFAR10 데이터셋이며, Y 는 클래스를 불균형하게 재구성한 CIFAR10 데이터셋에 생성된 이미지가 추가된 데이터셋이다. 본 논문은 기존 CIFAR10 데이터셋과 재구성된 데이터셋을 비교하기 위해 FID를 계산하기 위해 사전 학습된 Inception V3 모델을 사용하였다.

4.2.2 이미지 분류 성능 평가 지표

이미지 분류 성능의 평가 척도로는 정확도(Accuracy)와 F1-score를 사용하였다. 정확도는 이미지 분류 성능을 측정하기 위해 가장 기본적이고 대표적인 평가 지표며, F1-score는 정밀도(Precision)과 재현율(Recall)의 조화 평균으로 계산되는 통계적 지표이다. 정밀도는 모델이 양성으로 예측한 것 중에서 실제 양성인 것의 비율을 의미하며, 재현율은 실제 양성 중에서 모델이 양성으로 올바르게 예측한 것의 비율을 의미한다. F1-score는 모델의 정밀도와

재현율이 모두 높을 때만 큰 값을 갖기 때문에 클래스 불균형 문제에서 중요한 지표로 활용된다. 정확도와 F1-score를 계산하는 수식은 식 (12)과 같다.

$$\begin{aligned} Accuracy &= \frac{True}{True + False} \\ F1-score &= \frac{2 \times (Precision \times Recall)}{Precision + Recall} \\ Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \end{aligned} \quad (12)$$

클래스 불균형 데이터셋과 생성된 데이터셋이 합쳐져 학습된 모델을 통해, 기존의 CIFAR10 테스트셋 데이터에 대한 분류 성능이 검증되었다. 표 3에서는 생성 기법을 통해 클래스 불균형이 개선된 데이터로 학습한 모델의 정확도(Accuracy)를 비교하였다.

4.3 실험결과

4.3.1 이미지 생성 성능 평가

표 2은 대표적인 데이터 증강기법인 Mixup[19]와 CutMix[20], 그리고 본 연구에서 제안하는 두 가지 기법을 사용하여 생성된 이미지들의 FID 결과를 보여준다. CutMix의 이미지를 섞는 비율을 나타내는 하이퍼 파라미터는 0.6으로 설정하였다. 표 2의 첫 번째 행과 두 번째 행은 각각 기존의 데이터 증강 기법인 Mixup와 CutMix를 이용하여 생성된 이미지의 FID 결과를 나타낸다. 첫 번째로 제안된 기법인 CFG는 guidance scale(w)에 따라 다양한 이미지를 생성하는데, 본 논문에서는 guidance scale(w)에 따라 이미지 생성 품질을 비교하기 위해 두 가지 guidance scale(w)을 사용하여 이미지를 생성하였다. guidance scale(w)의 영향을 거의 받지 않는 $w = 0.1$ 와 LDM에서 사용한 guidance scale인 $w = 7.5$ 을 사용하여 비교하였다. 먼저, $w = 0.1$ 일 때의 FID는 28.028였으며 $w = 7.5$ 일 때는 FID가 17.678로 guidance scale(w)의 크기에 따라 생성된 이미지의 분포에 차이가 발생하며 guidance scale(w)이 클수록 기존 CIFAR10에 유사한 이미지를 생성함을 확인할 수 있었다.

또한, guidance scale(w)에 관계없이 CFG를 사용하여 재구성한 데이터셋 모두 기존의 Mixup, CutMix보다 기존 원본 CIFAR10 데이터셋의 분포와 유사한 것을 확인하였다.

표 2. 생성된 이미지가 추가되어 재구성된 데이터셋의 FID 성능 지표

Table 2. Performance of FID of reconstructed CIFAR10 datasets with generated images added

Method	FID
Mixup[19]	39.031
CutMix[20]	43.017
CFG($w = 0.1$)	28.028
CFG($w = 7.5$)	17.678
LDM	6.564

두 번째 제안 방법론인 LDM은 텍스트 프롬프트를 이용하여 각 클래스에 맞는 이미지를 생성하였다. 이 방법으로 생성된 이미지의 FID 값은 6.654로 기존 증강기법뿐 아니라 CFG보다도 더 높은 성능을 보여주었다.

그림 4는 제안된 생성 기법을 통해 생성된 이미지를 정성적으로 비교한 것이다. CFG의 경우 크기가 커질수록 이미지의 객체가 선명하게 나타났으며, LDM의 경우 CFG로 생성된 이미지보다 고품질의 이미지를 생성하며 ‘비행기’에 가까운 이미지를 생성하는 것을 확인할 수 있다.



그림 4. 제안된 방법론을 통해 생성된 이미지 결과
Fig. 4. Image generation results with proposed methods

4.3.2 이미지 분류 성능 평가

본 논문에서는 이미지 분류 성능을 평가하기 위해 ResNet18, ResNet50[18] 모델을 사용하였다. ResNet은 이미지 분류에 광범위하게 활용되는 대표적인 모델 중 하나이다. ViT[31], ConvNext[32] 등과 같이 최근에 제안된 더 깊고 복잡한 모델들은

CIFAR10을 적용할 경우, 데이터의 양이 상대적으로 적기 때문에 성능이 떨어지는 경향이 있어 선택하지 않았다. 이에, ResNet[18]만을 활용하여 분류 성능을 평가하였다.

ResNet18을 사용하여 기존의 데이터 증강기법인 Mixup[19]과 CutMix[20]로 재구성된 데이터를 학습한 경우에 비해, CFG($w = 0.1$)로 생성된 데이터를 적용하였을 때 정확도가 향상되었으며, 특히 CFG($w = 7.5$)로 생성된 데이터를 적용한 것은 더 높은 정확도를 보였다.

표 3. 생성된 이미지가 추가되어 재구성된 데이터셋의 정확도

Table 3. Performance of accuracy of reconstructed CIFAR10 with generated images added

Method	ResNet18	ResNet50
Mixup[14]	77.72	78.23
CutMix[15]	80.26	80.54
CFG($w = 0.1$)	79.33	80.28
CFG($w = 7.5$)	81.52	80.33
LDM	83.65	83.73

ResNet50에서도 기존 증강기법들보다 개선된 성능을 확인할 수 있었으나 ResNet18보다 성능이 개선되진 않았다. 이는 실험에 활용된 CIFAR10 데이터의 수가 많지 않으므로 네트워크의 깊이와 복잡성이 성능 향상을 보장하지 않는 것을 알 수 있다.

표 4. 생성된 이미지가 추가되어 재구성된 데이터셋의 F1-score

Table 4. Performance of F1-score of reconstructed CIFAR10 with generated images added

Method Class index	Mixup	CutMix	CFG $w = 0.1$	CFG $w = 7.5$	LDM
0	0.865	0.870	0.891	0.892	0.898
1	0.937	0.934	0.947	0.951	0.954
8	0.921	0.922	0.932	0.933	0.942
9	0.934	0.932	0.933	0.943	0.944

표 4는 각 클래스별 F1 score를 비교한 결과이다. 재구성된 실험 데이터셋에서 생성된 데이터의 비율이 높지 않은 클래스 8, 9번은 제안된 방법론과 기

존의 증강기법과 큰 차이가 없는 것을 확인할 수 있다. 반면, 생성된 데이터의 비율이 높은 클래스 0, 1번에서는 제안된 방법론의 F1 score가 더 높은 것을 확인할 수 있었다. 이를 통해서 제안된 방법론을 통해 재구성된 데이터셋이 불균형 문제를 효과적으로 완화되고 분류 모델이 소수 클래스를 더 잘 인식하게 되었다는 것을 알 수 있다.

또한, LDM의 분류 성능이 CFG를 포함한 기존 증강 기법보다 우수한 성능을 보이며 고해상도의 이미지 생성을 통한 데이터 증강이 분류 성능의 향상에 유의미한 영향을 끼쳤음을 확인하였다.

V. 결론 및 향후 과제

본 논문에서는 클래스 불균형 이미지 데이터에 분류 성능 개선을 위한 데이터 증강기법으로 Diffusion 모델 기반 2가지 방법론을 제시하였다. 첫 번째로 제안한 방법론은 클래스 불균형 데이터셋을 학습한 Classifier-free guidance를 활용해 소수 클래스의 이미지를 생성하였다. 두 번째로 제안한 방법론은 사전 학습된 LDM의 텍스트 입력을 클래스명으로 사용해 소수 클래스의 이미지를 생성하였다. 제안된 방법론을 통해 생성된 데이터셋은 클래스 불균형 데이터셋과 결합해 재구성하여 이미지 분류 모델인 ResNet으로 학습하였다.

제안된 두 방법론은 이미지 생성 및 분류 분야에서 널리 활용된 기존 데이터 증강 기법들에 비해 효과적인 성능을 보였으며, 특히 두 번째로 제안된 방법론인 LDM을 활용한 기법이 가장 우수한 성능을 보였다. 두 가지 제안된 방법론 모두 기존 증강 기법에 비해 원본 CIFAR10 데이터셋과 유사한 이미지를 생성하며 데이터를 증강하였고, 이는 이미지 분류 성능의 향상으로 이어졌다.

본 연구의 한계점 중 하나는 CFG가 소수 클래스의 이미지를 효과적으로 학습하고 생성하는 데 성공했지만 데이터셋의 크기가 더욱 줄어들 경우 일관된 성능을 유지하는 데 어려움이 있다는 것이다. LDM의 경우 사전 학습된 모델을 활용하여 적은 클래스 이미지를 고품질로 생성할 수 있다. 그러나 실험에 사용된 데이터는 비교적 간단한 이미지(예:

개, 고양이)로 구성되어 있어 사전 학습된 LDM으로 고품질의 이미지를 효과적으로 생성할 수 있었지만, 실제 산업에서의 복잡하고 희소한 데이터를 생성하는 것에는 한계점이 존재한다. 따라서, 실제 산업에서 클래스 불균형 문제를 해결하려면 기존 모델에 관련 이미지를 추가로 학습하는 등의 작업을 고려할 수 있다. 본 연구에서 제안하는 생성 기반 분류 모델은 클래스 불균형 문제를 가진 다양한 분야에서의 확장된 연구와 실제 활용 가능한 방법으로서의 적용을 기대해 볼 수 있다.

References

- [1] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network", 2014 13th international conference on control automation robotics & vision (ICARCV), Singapore, Dec. 2014. <https://doi.org/10.1109/ICARCV.2014.7064414>.
- [2] H. Wang, Q. Zhang, H. Lu, D. Won, and S. W. Yoon, "3D medical image classification with depthwise separable networks", *Procedia Manufacturing*, Vol. 39, pp. 349-356, 2019. <https://doi.org/10.1016/j.promfg.2020.01.369>.
- [3] G.-E. Kim and C.-H. Son, "Multiscale Crosss Attention Vision Transformer for Pest Image Classification", *The Journal of Korean Institute of Information Technology*, Vol. 21, No. 7, pp. 77-84, Jul. 2023. <https://doi.org/10.14801/jkiit.2023.21.7.77>.
- [4] K. Imoto, T. Nakai, T. Ike, K. Haruki, and Y. Sato, "A CNN-Based transfer learning method for defect classification in semiconductor manufacturing", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 32, No. 4, pp. 455-459, Nov. 2019. <https://doi.org/10.1109/TSM.2019.2941752>.
- [5] J. Myeong, "Classify products in shopping malls based on deep learning models using image and text information", *Korean Institute of Information*

- Scientists and Engineers, pp. 681-683, Dec. 2020.
- [6] E. Choi and K. Yoon, "Fashion Image Classification System Considering Object Aspect Ratio", *Journal of Broadcast Engineering*, Vol. 29, No. 1, pp. 1-9, Jan. 2024. <https://doi.org/10.5909/JBE.2024.29.1.1>.
- [7] Y. Sho, M. Son, H. Park, S. Yoon, H. Yi, and I. Yoon, "Image Classification of Insects of Rice Using Machine Learning", *Proc. of KIIT Conference*, Jeju, Korea, pp. 466-467, Jun. 2022.
- [8] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning", *arXiv preprint arXiv:1712.04621*, pp. 1-8, Dec. 2017. <https://doi.org/10.48550/arXiv.1712.04621>.
- [9] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes", *arXiv preprint arXiv:1312.6114*, Dec. 2013. <https://doi.org/10.48550/arXiv.1312.6114>.
- [10] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview", *IEEE signal processing magazine*, Vol. 35, No. 1, pp. 53-65, Jan. 2018. <https://doi.org/10.1109/MSP.2017.2765202>.
- [11] D.-G. Kang and J. Choi, "Validation of Generative Adversarial Network Based Generated images in Class Imbalance Situation", *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 28, No. 2, pp. 139-145, Feb. 2024. <https://doi.org/10.6109/jkiice.2024.28.2.139>
- [12] W. Liu and W. Hu, "Oversampling of Tabular Data for Imbalanced Learning Via Denoising Diffusion Probabilistic Models", Available at SSRN 4673719, Dec. 2023. <http://dx.doi.org/10.2139/ssrn.4673719>.
- [13] Q. Ai, P. Wang, L. He, L. Wen, L. Pan, and Z. Xu, "Generative Oversampling for Imbalanced Data via Majority-Guided VAE", *arXiv preprint arXiv:2302.10910*. Feb. 2023. <https://doi.org/10.48550/arXiv.2302.10910>.
- [14] A. Majeed and S. O. Hwang, "CTGAN-MOS: Conditional Generative Adversarial Network Based Minority-Class-Augmented Oversampling Scheme for Imbalanced Problems", *IEEE Access*, Vol. 11, pp. 85878-85899, Jun. 2023. <https://doi.org/10.1109/ACCESS.2023.3303509>.
- [15] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep Unsupervised Learning using Nonequilibrium Thermodynamics", In *International conference on machine learning*, pp. 2256-2265, PMLR, Nov. 2015. <https://doi.org/10.48550/arXiv.1503.03585>.
- [16] J. Ho and T. Salimans, "Classifier-Free Diffusion Guidance", *arXiv preprint arXiv:2207.12598*, Jul. 2022. <https://doi.org/10.48550/arXiv.2207.12598>.
- [17] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models", In *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684-10695, Dec. 2021. <https://doi.org/10.48550/arXiv.2112.10752>.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", In *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, Dec. 2015. <https://doi.org/10.48550/arXiv.1512.03385>.
- [19] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond Empirical Risk Minimization", *arXiv preprint arXiv:1710.09412*, Oct. 2017. <https://doi.org/10.48550/arXiv.1710.09412>.
- [20] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features", In *Proc. of the IEEE/CVF international conference on computer vision*, pp. 6023-6032, May 2019. <https://doi.org/10.48550/arXiv.1905.04899>.
- [21] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, "A GAN-based Data Augmentation Method for Imbalanced Multi-class Skin Lesion

- Classification", IEEE Access, Vol. 12, pp. 16398-16513, Jan. 2024. <https://doi.org/10.1109/ACCESS.2024.3360215>.
- [22] S. J. Bang, M. J. Kang, M. G. Lee, and S. M. Lee, "STO-CVAE: state transition-oriented conditional variational autoencoder for data augmentation in disability classification", Complex Intelligent System, Mar. 2024. <https://doi.org/10.1007/s40747-024-01370-x>.
- [23] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models", Advances in neural information processing systems, Vol. 33, pp. 6840-6851, Jun. 2020. <https://doi.org/10.48550/arXiv.2006.11239>.
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, Vol. 9351, pp. 234-241, May 2015. https://doi.org/10.1007/978-3-319-24574-4_28.
- [25] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis", Advances in neural information processing systems, Vol. 34, 8780-8794, May 2021. <https://doi.org/10.48550/arXiv.2105.05233>.
- [26] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, and M. Chen, "GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models", arXiv preprint arXiv:2112.10741, Dec. 2021. <https://doi.org/10.48550/arXiv.2112.10741>.
- [27] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, and M. Norouzi, "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding", Advances in Neural Information Processing Systems, Vol. 35, pp. 36479-36494, May 2022. <https://doi.org/10.48550/arXiv.2205.11487>
- [28] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical Text-Conditional Image Generation with CLIP Latents", arXiv preprint arXiv:2204.06125, Apr. 2022. <https://doi.org/10.48550/arXiv.2204.06125>.
- [29] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", Advances in neural information processing systems, Vol. 30, Jun. 2017. <https://doi.org/10.48550/arXiv.1706.08500>.
- [30] J.-C. Park, J.-U. Heu, and G.-W. Kim, "Enhancing Rebar Endpoint Detection Performance through High-Resolution Rebar Image Generation with SRPAttention-GAN", The Journal of Korean Institute of Information Technology, Vol. 21, No. 11, pp. 39-52, Nov. 2023. <https://doi.org/10.14801/jkiit.2023.21.11.39>.
- [31] H. Chang, J.-S. Lee, and J.-H. Lee, "Aleevitating Class Imbalance Problems based on Mixup-based Noise Sampling of Diffusion Generative Model", Journal of Korean Institute of Intelligent Systems, Vol. 33, No. 3, pp. 222-227, Jun. 2023. <http://dx.doi.org/10.5391/JKIIS.2023.33.3.222>.
- [32] A. Dosovitskiy, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", arXiv preprint arXiv:2010.11929, Oct. 2020. <https://doi.org/10.48550/arXiv.2010.11929>.
- [33] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11976-11986, Jan. 2022. <https://doi.org/10.48550/arXiv.2201.03545>.

저자소개

유 대 현 (Daehyun Yoo)



2022년 8월 : 연세대학교
정보통계학과(학사)
2022년 9월 ~ 현재 : 고려대학교
산업경영공학과 석사과정
관심분야 : 인공지능, 컴퓨터비전,
이미지 생성

김 지 호 (Jiho Kim)



2015년 8월 : 서울과학기술대학교
글로벌융합산업공학과(학사)
2015년 9월 ~ 현재 : 고려대학교
산업경영공학부 석박통합과정
관심분야 : 인공지능, 자연어처리,
비즈니스 인텔리전스

이 흥 철 (Hongchul Lee)



1983년 : 고려대학교
산업경영공학부(학사)
1986년 ~ 1988년 : University of
Texas Arlington, Industrial
Engineering (M.S.)
1988년 ~ 1993년 : Texas A&M
University, Industrial

Engineering (Ph D.)

1995년 ~ 1996년 : Research Committee, Production and
technology Institute of Korea University

1993년 ~ 1994년 : The University of Iowa, Post Doc

1986년 ~ 1993년 : Research and Teaching Assistant,
Texas A&M University

1984년 ~ 1986년 : Mathematics lecturer Korea Military
Academy

1996년 ~ 현재 : 고려대학교 산업경영공학부 교수

관심분야 : 인공지능, 생산공학시스템, 시뮬레이션