전처리 방식에 따른 AI 이미지와 실사 이미지의 분류 성능 비교

권소희[a] and 최현성[b], 김웅기[b], 엄소은[b], 손명균[b]

계명대학교 공과대학
Tel: 053-580-6361, E-mail: efgihj@nyu.edu

계명대학교 경영대학
Tel: 042-000-0000, E-mail:abcde@yonsei.ac.kr

계명대학교 경영대학
Tel: 053-580-6361, E-mail: efgihj@nyu.edu

계명대학교 공과대학
Tel: 053-580-6361, E-mail: efgihj@nyu.edu

계명대학교 자연과학대학
Tel: 053-580-6361, E-mail: efgihj@nyu.edu

요약

본 연구는 생성이미지와 실사이미지의 분류를 위한 최적의 이미지 전처리 기법을 탐색한다. Stable diffusion과 크롤링을 통해 구축한 데이터셋에 sharpening, grayscale, canny, diffusion, ELA(Error Level Analysis) 총 5가지 전처리 기법을 적용하여 모델별 성능을 비교하고자 한다.

Keywords

Image Preprocessing, AI Generated image, CNN, ViT, Computer Vision, Deep Learning, Stable Diffusion, GrayScale, Sharpening, Canny, ELA, Deep Fake

I. 서론

우리는 인공지능의 발전에 따라 자율주행, 챗봇, 음성인식 인공지능 서비스 등 일상생활에서 해당 기술의 편리함을 누리며 살아가고 있다. 인공지능은 현대 딥러닝 기술의 발전과 컴퓨터 비전 분야의 혁신적인 연구로, 이미지 생성 분야에서도 큰 발전을 이루었다. 다양한 생성 모델들은 현실적으로 보이는 이미지를 생성하고, 이는 컴퓨터 비전뿐만 아니라 예술, 경영, 시뮬레이션 연구 등 여러 분야에서 활용되고 있다. 그러나 이러한 발전이 가져올 변화는 마냥 긍정적이지 않다. 더 정교하고 더 현실적인 AI 이미지 생성 모델의 발전으로 생긴 주요 문제 중 하나는 해당 모델이 생성한 이미지와 실사 이미지 간의 경계가 뚜렷하지 않아 구분하기 어려워진다는 것이다. 따라서 본 연구의 배경은 딥러닝 모델이 발전하면서 AI가 생성한 이미지가 정교하고 사실적으로 구현되었으며, 이로 인해 이미지의 진위성을 판단하기 어려워지면서 이미지와 영상에 대한 신뢰성과 진실성을 검증하는 데 중요한 역할을 하는 이미지 분류 및 인증의 어려움이 더욱 심각해지고 있다는 것에 기인되었다.

AI 이미지 생성 모델의 발전으로 인해 실제 이미지와 AI가 생성한 이미지 간의 구분이 점점 어려워지면서 생길 수 있는 가장 큰 문제점은 정보의 신뢰성 저하이다. AI가 현실적인 이미지 구현이 가능해지면서, 실사 이미지와의 차이를 식별하기 어려워졌다. 이의 예시로 딥페이크(Deepfake)를 들 수 있다. 딥페이크는 deep learning과 fake의 합성어로 딥러닝과 인공 지능 기술을 활용하여 사람의 얼굴, 음성 또는 다른 특징을 모방하거나 합성하여 실제와 구별하기 어려운 가짜 동영상, 음성, 이미지 등을 생성하는 기술을 가르킨다. 인공지능 기술이 발달하는 과정에서 딥페이크는 고도로 현실적인 콘텐츠를 모방하는 것을 가능하게 하였고 매우 진보된 결과를 도출해냈다. 실제로 이는 놀랍고 혁신적인 기술일 수 있지만, 이러한 기술은 악의적인 목적으로 사용될 때 심각한 문제를 일으킬 수 있다. 딥페이크로 만들어진 콘텐츠가 가짜 정보나 왜곡된 내용을 전파하는 데 악용될 수 있으며, 이는 공중에 유포된 정보와 데이터의 신뢰성을 훼손한다. 더 나아가, 이러한 가짜 콘텐츠는 사회적, 윤리적, 그리고 법적 문제를 일으키며, 명예 훼손, 개인 정보 침해, 대중에게 혼란을 초래할 수

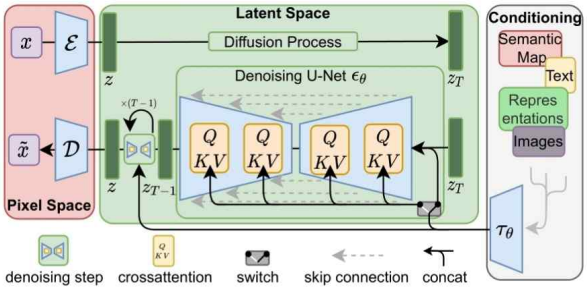
있다.
본 연구의 목적은 이러한 문제들을 이미지 분류 분야에서 AI 생성 이미지와 실제 이미지를 구별하는데 도움을 주는데 있으며 최적의 전처리접근 방식을 제시하고, 다양한 딥 러닝 모델의 성능을 평가함으로써 현대 사회의 미디어 정보의 신뢰성과 진실성을 높이며 이 분야에 대한 근거를 제시하고자 한다.

II. 사전연구

1. Stable Diffusion

Stable Diffusion은 사전 훈련된 AutoEncoder의 latent space에서 훈련하여 확산 모델의 계산 요구 사항을 완화한다. 이 접근 방식은 최적화 비용과 추론 복잡성을 줄이면서 품질과 유연성을 유지한다. Stable Diffusion은 모델 아키텍처에 교차 어텐션 레이어의 사용을 제안하여 텍스트나 BoundingBox와 같은 조건 입력을 가능하게 한다. 또한,입력 이미지에 noise를 추가하여 denoising step을 수행하여 학습의 성능을 올릴수있고, denoising U-Net을 사용하여 이미지의 노이즈를 제거하여 더 정밀한 이미지 생성을 가능하게 한다. 해당모델은 이미지 인페인팅, 클래스 조건 이미지 생성, 텍스트에서 이미지 생성, 무조건 이미지 생성 및 초해상도와 같은 다양한 작업에서 최신 기술 수준의 성능을 낸다. 픽셀 기반 확산 모델과 비교했을 때 Stable Diffusion은 계산 요구 사항을 크게 줄인다

<Figure 1> Stable Diffusion Architecture



2. Canny

세트를 수학적으로 표현하여 검출자의 동작을 명확히 한다. 이 알고리즘은 엣지의 검출과 위치 지정 기준을 정의하고, 이를 연산자의 임펄스 응답에 대한 수학적인 형태로 표현한다. 또한, 한 개의 엣지에 대해서만 한 개의 응답을 갖도록 하는 추가적인 기준도 제시한다. 이 알고리즘은 다양한 이미지 특징을 위한 검출자를 구현하기 위해 수치 최적화를 사용한다. 또한, 스텝 엣지에 특화된 분석을 통해 검출과 위치 지정 성능 사이에 자연스러운 불확정성 원리가 있다는 것을 발견하며, 이를 고려하여 어떤 스케일에서도 최적의 검출자를 유도한다. 최적의 검출자는 Gaussian-smoothed 이미지에서 gradient 크기의 최댓값을 검출하여 엣지를 표시하는 간단한 근사 구현을 갖는다. 이 알고리즘은 다양한 신호 대 잡음 비율에 대응하기 위해 여러 폭의 연산자를 사용하여 확장할 수 있다. 또한, 연산자의 포인트 스프레드 함수가 엣지를 따라 확장됨에 따라 스텝 엣지 검출자의 성능이 크게 향상됨을 보여준다

3. GrayScale

grayscale은 낮은 대비 및 음영 효과로 인해 객체 추출이 어려운 상황에서 사용된다. 이 기술은 입력 이미지를 두 개의 통합기로 분할하고, 초기 추정값과 스위칭 함수를 사용하여 이미지 신호를 배경과 객체로 분류한다. 임계값은 통합기의 출력을 평균하여 결정되며, 반복적으로 수행하여 점점 더 깨끗한 객체 영역을 추출한다. 예시에서 보여지는 것처럼 초기 추정값은 배경 잡음을 제거하지 못하지만, 반복이 진행됨에 따라 점점 더 정확한 추출이 가능해진다. 이를 통해 grayscale 이미지에서 높은 품질의 객체 추출이 깨끗해진다.

4. Sharpening

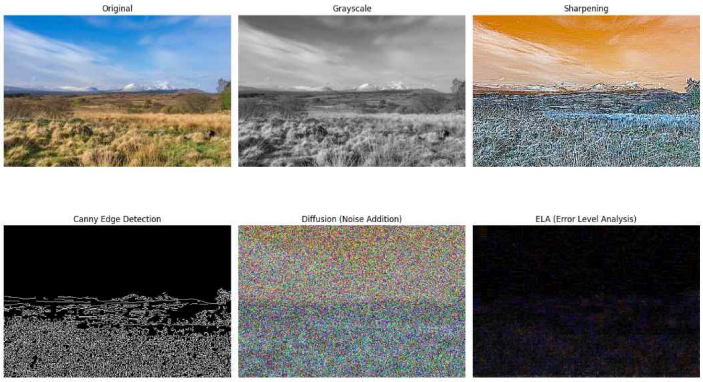
엣지 정보는 이미지의 특정 부분에서 밝기가 급격하게 변화하는 부분을 나타내며, 샤프닝은 이러한 엣지를 검출하고 강화하는 작업을 수행한다. 엣지 검출은 이미지에서 짧은 범위에서 밝기의 급격한 변화를 식별하는 과정으로, 이미지 처리에서 중요한 역할을 한다. 엣지 정보를 검출한 후, 해당 부분에 이미지의 상세한 정보를 강화시키는 연산을 수행함으로써 물체의 경계나 질감의 경계를 선명하게 만들어준다. 이미지 처리에서 엣지 정보는 중요한 시각적 특징으로, 샤프닝과 같은 기술은 엣지 정보를 뚜렷하게 표현하고 이미지의 질을 향상시키는 데에 활용된다. 샤프닝은 이미지 내의 엣지를 검출하고 강화하는 방법 중 하나로, 이미지 처리 분야에서 폭넓게 활용되고 있다. 이러한 기술을 통해 이미지의 시각적 품질을 향상시키고 다양한 응용 분야에서 더 나은 결과를 얻을 수 있다.

III. 실험방법

1. 실험준비

본 연구에서는 생성 이미지와 실사 이미지간의 분류 성능을 비교하기 위해, 총 5가지 전처리 기법과 4개

의 모델을 사용하였다. 데이터는 Stable diffusion을 통해 생성한 3497개의 생성 이미지와 웹 크롤링을 통해 수집한 4925개의 실사 이미지로 데이터셋을 구축하였다. 모든 이미지는 224*224로 리사이징(Resizing)하였으며, ImageNet이 학습한 수백만장의 이미지에서 추출한 RGB 각각의 채널에 대한 평균 0.485, 0.456, 0.406, 표준편차 0.229, 0.224, 0.225로 정규화해주었다. 그리고 이미지의 특성을 변형시켜 모델이 더 강력하고 강인한 특징을 학습할 수 있도록 돕기 위해 Grayscale, Canny, Sharpening, ELA(error level analysis), Diffusion 총 5가지의 전처리를 수행하였다.



<Figure 2> Image Preprocessing

모델은 Vision Transformer (ViT), Resnet, EfficientNet, CoAtNet 총 4가지 모델을 사용하였다. 각 모델은 별도의 파라미터 수정 없이, batch_size=64로 총 10 epochs 동안 학습을 진행하였다. 학습 과정에서 각 epoch마다 모델의 성능을 평가하고, 가장 높은 정확도를 달성한 모델을 저장하였다.

2. 결과해석

총 5개의 데이터셋, 4개의 모델을 통해 20개의 모델을 구축하였으며, 상세한 성능 비교를 위해 추가적으로 학습에 사용되지 않은 310개의 Stable diffusion 생성 이미지와 670개의 실사 이미지를 수집하였다. 전체적인 정확도 파악을 위한 accuracy와 생성 이미지의 오탐지율 파악을 위한 recall을 중점적으로 살펴보았다.

<Table 1> Model Performance Results

		Acc	Recall
Grayscale	ViT	0.673	0.777
	Resnet	0.707	0.948
	EfficientNet	0.902	0.796
	CoAtNet	0.201	0.635
Canny	ViT	0.683	0.000
	Resnet	0.683	0.000
	EfficientNet	0.765	0.258
	CoAtNet	0.890	1.000
Sharpening	ViT	0.627	0.309
	Resnet	0.768	0.280
	EfficientNet	0.922	0.767
	CoAtNet	0.820	0.438
ELA	ViT	0.683	0.000
	Resnet	0.414	1.000
	EfficientNet	0.995	0.996
	CoAtNet	0.728	0.141
Diffusion	ViT	0.683	0.000
	Resnet	0.997	0.993
	EfficientNet	0.808	0.393
	CoAtNet	0.316	1.000

전처리별 모델 성능 비교를 통해 각각의 특징을 분석했다. ViT 모델에서는 대부분 좋지 않은 성능을 보였다. 이는 데이터 셋의 크기가 커질수록 성능이 좋아지는 ViT의 특성상 학습에 사용된 데이터의 양이 작기에 나타나는 문제라고 판단하였다. Canny와 Diffusion, ELA와 같이 원본의 특성을 극한으로 추출하는 전처리 기법에서는 과적합, 과소적합이 의심되는 모델 성능을 보인다. 특히, Accuracy<0.5 이거나 Recall=0인 극단적인 성능을 보이는 모델들의 각 예측에 대한 확률(Probability)를 확인해본 결과, 이미지에 상관없이 50~60% 확률을 보이고 있다. 모델이 이미지의 특징을 정상적으로 학습하지 못해, 불확실한 예측확률을 보이는 것이라 판단하였다. 반면, Grayscale, Sharpening과 같이 원본을 최대한 보존하며 최소한의 특성을 추출하는 전처리 기법에서는 정확하고 정상적인 모델 성능을 보인다. 그 중 Resnet과 EfficientNet의 성능이 높아 CNN 기반 모델이 생성이미지와 실사이미지 분류에 있어서 가장 뛰어나다고 분석하였다.

IV. 결 론

본 논문에서는 Stable diffusion을 이용하여 풍경 이미지를 생성하여 실사 이미지와의 분류를 위한 모델을 구축하였다. 연구 결과 성능 측면에서 Grayscale, Sharpening과 같이 최소한의 특성을 추출하는 전처리 기법이 가장 효과적이었으며, CNN 기반의 모델을 사용하는 것이 가장 좋은 성능을 나타내었다. 본 논문에서는 적은 데이터셋 분량과 Stable diffusion을

통해 생성한 이미지만을 사용하였다는 한계점이 존재한다. 향후 다양한 이미지 생성 모델을 사용하여 데이터셋의 크기와 다양성을 늘린다면 더욱 정확하고 범용성 있는 연구가 될 것으로 기대된다.

참고문헌

[1] John Canny(1986). A Computational Approach to Edge Detection

[2] T. W. RIDLER AND S. CALVARD(1978). Picture Thresholding Using an Iterative Selection Method

[3] Jonnadula Narasimharao(2023). Digital image processing

[4]Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. CVPR 2022.

[4] Rombach, R., Blattmann, A., Lorenz, D., ESSER, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. CVPR 2022.