인공지능 기반 스테가노그래피 생성 기술 최신 연구 동향

(Research Trends in Steganography Based on Artificial Intelligence)

김현지*, 임세진*, 김덕영*, 윤세영*, 서화정**
(Hyun Ji Kim, Se Jin Lim, Duk Young Kim, Se Young Yoon, Hwa Jeong Seo)

요 약

스테가노그래피는 데이터의 존재 자체를 은닉하여 데이터를 보호하는 기술이다. 최근에는 인공지능 기술이 발달함에 따라 딥러닝 기반의 스테가노그래피 기법들이 개발되고 있다. 딥러닝 기술은 데이터에 대한 고차원의 특징을 분석하여 학습할 수 있으므로 스테가노그래피의 성능과 품질을 개선시킬 수 있다. 본 논문에서는 이미지 데이터에 대한 딥러닝 기반의 스테가노그래피 기술의 최신 연구 동향에 대해 살펴보도록 한다.

■ 중심어 : 인공지능 ; 딥러닝 ; 스테가노그래피 ; 정보 보호

Abstract

Steganography is a technology capable of protecting data by hiding the existence of data. Recently, with the development of deep learning technology, deep learning-based steganography are being developed. Deep learning can learn by analyzing high-dimensional features of data, so it can improve the performance and quality of steganography. In this paper, we investigated the research trend of image steganography based on deep learning.

■ keywords : Artificial Intelligence ; Deep learning ; Steganography ; Data security

Ⅰ. 서 론

최근 빅데이터 산업들의 발전으로 인해 데이터의 양이 증가하고 있으며, 중요 정보들을 보호하기 위해 스테가노그래피 (Steganography) 기술이 활용되고 있다[1,2]. 스테가노그래피는 이미지 등의 매체에 비밀 정보를 임베딩하고 그 존재 자체를 숨기는 것이 목적인 데이터 은닉기술이다. 최근에는 인공지능 기술의 발달로 인해 데이터에 대한 고도화된 분석이 가능해짐에 따라 딥러닝 기반의 정교한 스테가노그래피 기술들이 등장하고 있다. 본 논문에서는 이미지 데이터에 대한 딥러닝 기반 스테가노그래피 기술의 최신 연구 동향에 대해 살펴본다.

Ⅱ. 본 론

표1은 본 논문에서 소개되는 스테가노그래피 기술들을 간단히 정리한 표이다. 크게 고전 방식과 딥러닝 기반의 방식으로 나뉘고, 각 방식에는 기본적인 기술 (LSB, PVD, CNN)과 조금 더 고도화된 기술들 (WOW, UNIWARD, GAN)이 존재한다.

표 1. Steganography technologies by method

Method	Technologies
Classical steganography	LSB
	PVD
	WOW
	UNIWARD
	DCT
Deep learning based steganography	CNN
	GAN

1. 고전적인 스테가노그래피

스테가노그래피는 중요 정보를 숨기기 위해 텍스트, 이미지, 동영상 등의 미디어 파일에 데이터를 삽

* 학생회원, 한성대학교 IT융합공학과
** 종신회원, 한성대학교 융합보안학과
This research was financially supported by Hansung University.
접수일자 : 2023년 04월 12일
수정일자 : 1차 2023년 05월 18일
게재확정일 : 2023년 05월 24일
교신저자 : 서화정 e-mail : hwajeong84@gmail.com

입하는 기술이다. 이때, 비밀 데이터는 눈에 띄지 않도록 삽입되며, 기존의 파일과 동일한 모습을 가진다. 스테가노그래피는 데이터 암호화와 임베딩 과정으로 구성된다. 전체 과정은 그림 1과 같으며, 먼저 비밀 데이터(M)를 암호화 한다. 이때, 사용된 암호화 종류에 따라 크게 일반, 비밀키, 공개키 스테가노그래피로 나뉜다. 일반 스테가노그래피는 비밀 데이터가 암호화되지 않은 채로 임베딩 되기 때문에 임베딩 알고리즘의 성능에만 의존하게 되어 보안성이 다소 낮다. 비밀키와 공개키 스테가노그래피는 각각 대칭키 및 공개키 암호화를 통해 비밀 메시지에 대한 기밀성을 확보한다. 이 두 방식은 일반 스테가노그래피의 보안성의 한계를 극복할 수 있으나, 암호화로 인해 비밀 데이터의 크기가 증가할 수 있다. 이로 인해 숨길 수 있는 비밀 데이터의 크기가 줄어들고, 커버 오브젝트 (비밀 데이터를 숨기고자 하는 대상이 되는 매체)의 용량이 증가함에 따라 스테가노그래피의 존재 여부가 발각되기 쉽다. 이러한 단점을 보완하기 위해 비트별 암호화인 스트림 암호 (RC4)가 적용된 연구 사례가 있다[3,4].

비밀 데이터를 암호화 한 후에는 비밀 데이터를 커버 오브젝트에 숨겨서 스테고 오브젝트를 생성하는 과정이 수행된다. 이러한 과정을 임베딩이라고 하며, 임베딩은 커버 오브젝트의 왜곡 및 손상을 최소화하여 비밀 정보가 은닉되었다는 사실을 숨길 수 있어야 한다. 즉, 스테가노그래피의 성능은 임베딩할 수 있는 용량 (숨길 수 있는 데이터의 양, 주로 픽셀 당 비트의 수 (bit per pixel, bpp)[5]), 왜곡 (커버와 스테고 간의 유사성), 보안성 (스테그아날리시스에 대한 저항성) 의해 평가된다. 일반적으로 비밀 데이터가 검출되기 어려워지는 위치에 비밀 데이터를 삽입하는 임베딩 방식이 더 높은 성능을 보인다. 대표적인 스테가노그래피 임베딩 방식들은 다음과 같다.

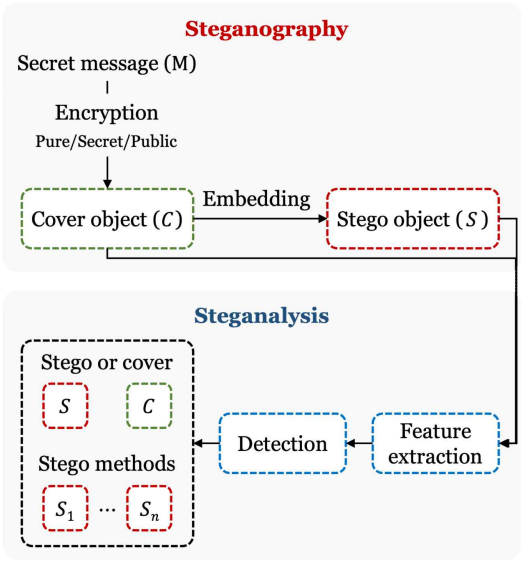

그림 1 스테가노그래피 및 스테그아날리시스

가. Least Significant Bit (LSB)

LSB 방식[6,7]은 픽셀의 최하위비트를 변조함으로써 비밀 데이터를 삽입한다. 예를 들어 픽셀의 값이 254인 경우, 이를 이진수로 표현하면 11111110이 되고, 여기에서 삽입하려는 데이터인 1을 임베딩하면 해당 픽셀의 값은 11111111이므로 255가 된다. LSB 방식은 n-LSB 방식으로 확장될 수 있으며, 이는 최하위 n개의 비트를 조작하는 방식이다. 이처럼 픽셀 값에 큰 차이가 없는 비트 단위의 조작이 이루어지므로 큰 왜곡이 발생하지 않는다. 그러나 비트 단위의 조작을 수행하기 때문에 숨길 수 있는 비밀 데이터의 양이 다른 방식들에 비해 상대적으로 적다. 또한, 임베딩 용량을 늘린다면 값의 차이가 큰 비트들까지 변경되어 이미지의 전반적인 품질이 저하된다.

나. Pixel Value Differencing (PVD)

PVD 방식[8,9]은 LSB 방식의 품질 저하 문제를 해결하기 위해 제안되었다. 인접한 두 픽셀의 차이 값에 따라 데이터를 삽입하는 방식으로 픽셀 간의 차이가 클수록 더 많은 데이터를 삽입할 수 있다. 이는 시각적으로 뚜렷하게 보이는 엣지 부분은 픽셀 값의 차이가 크므로 더 많은 데이터를 삽입하고, 유사한 픽셀 값을 갖는 부분에는 적은 양의 데이터를

삽입하도록 하는 방식이다. 이를 통해 시각적인 왜곡을 줄여 차이를 인지하기 어렵도록 한다.

전체적인 과정은 그림 2와 같다. 먼저 2개의 픽셀을 하나의 서브블록으로 하여 전체 이미지를 나눈다. 이후, 두 픽셀 간의 차이 값 (d)을 계산하여 삽입할 비밀 데이터의 크기 (n)를 결정한다. n은 두 픽셀의 차이 값을 여러 구간으로 나누어 둔 표를 기반으로 정해진다. 이때, 구간을 나누는 기준은 사람의 시각 감도를 기반으로 정해지며, 두 픽셀 간의 차이가 적을수록 더 적은 비트의 데이터를 삽입할 수 있다.

삽입할 데이터의 크기가 정해지면, 삽입하고자 하는 데이터에서 해당 크기만큼의 비트를 가져온 후, 이를 10진수로 변환한다. 10진수로 변환된 비밀 데이터 값과 차이 값에 의해 정해진 구간의 최솟값을 더하여 새로운 차이 값(d_n)을 정한다. 마지막으로 d_n을 2로 나눈 값을 기존의 픽셀 값에 더하거나 빼서 새로운 픽셀 값으로 대체하여 스테고 데이터를 생성한다. 그러나 일반적인 이미지에는 경계선보다 유사한 픽셀 값을 갖는 부분이 더 많다. PVD는 주로 인접 픽셀 간의 차이가 큰 경계선에 데이터를 은닉하므로 많은 데이터를 임베딩하기 어렵다. 또한, 히스토그램을 통해 분석할 경우 비밀 데이터의 존재가 발각되기 쉽다[10,11].

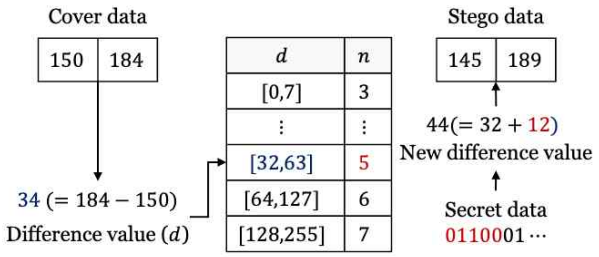

그림 2 PVD 방식

다. Wavelet Obtained Weights (WOW)

WOW는 LSB, PVD와 다르게 데이터의 삽입 용량 및 품질보다 데이터의 은닉성에 초점을 둔 임베딩 알고리즘이다. LSB와 PVD 방식은 이미지의 각 픽셀 간의 관계를 고려하기보다는 모든 픽셀에 대해 순차적으로 임베딩을 수행하였고, 이는 비밀 데이터가 탐지될 가능성을 높인다. 그러나 WOW는 주변 픽셀들을 고려하여 커버 데이터와 스테고 데이터가 유사한 분포를 가지도록 한다. 이를 위해 비밀 데이터를 임베딩할 때 발생하는 픽셀의 변화 비용을 계산하고, 해당 비용을 바탕으로 왜곡이 최소가 되도록 하는 픽셀에 데이터를 은닉한다[12].

전체적인 동작 과정은 다음과 같다. 우선, 커버 데이터에 대해 미러 패딩 (인접 픽셀의 값으로 패딩)을 적용한 후, 방향성을 갖는 필터와 커버 데이터 간의 컨볼루션 연산을 수행하여 잔차 (Residual)를 계산한다. 다음으로는 방향성 필터와 계산된 잔차를 컨볼루션하여 모든 픽셀과 방향성 필터에 대한 적합도를 구한다. 마지막으로 계산된 적합도의 역수를 모두 더하여 픽셀들의 변화 비용을 계산한다. 이러한 방식을 사용하면 인접 픽셀들과의 관계를 반영하여 데이터의 왜곡을 최소화하는 방향으로 임베딩할 수 있게 된다.

라. Universal Wavelet Relative Distortion (UNIWARD)

UNIWARD는 WOW를 개선한 임베딩 알고리즘이다. UNIWARD에서는 WOW에서 사용하는 필터 중 하나인 WDFB-B 필터를 사용하여 잔차를 계산한다. WOW와 UNIWARD 방식은 커버 데이터와 스테고 데이터의 픽셀 분포가 유사하기 때문에 비밀 데이터가 탐지될 확률이 낮은 강력한 임베딩 기술이다. 그러나 WOW에서 비용 계산에 사용되는 적합도를 주파수 도메인에서 그대로 사용할 경우, 탐지되기 쉬운 위치에 비밀 데이터를 삽입할 수 있는 문제점이 존재한다. 따라서 UNIWARD에서의 적합도는 WOW의 적합도의 역수로 정의하며 공간 도메인 뿐만 아니라 주파수 도메인에도 적용 가능한 범용적인 임베딩 기술으로 사용되고 있다[13]. 즉, 주파수 도메인에서 탐지되기 쉬운 WOW 알고리즘을 개선한 것이 UNIWARD이고, 스테그아날리시스에 더욱 강인하며 공간 및 주파수 도메인에서 사용할 수 있는 범용적인 임베딩 기술이다. UNIWARD에는 S-UNIWARD 및 J-UNIWARD가 있는데, 이는 각각 공간 및 주파수 도메인에서의 UNIWARD 알

고리즘을 의미한다.

마. Discrete Cosine Transform (DCT)

DCT (이산 코사인 변환)는 공간 도메인의 신호를 주파수 도메인으로 변환하는 기술이다[14]. 낮은 비트 오류율, 큰 압축비와 같은 장점으로 인해 이미지, 영상, 오디오 등 디지털 미디어 압축에 사용되고 있다[15]. DCT를 사용하여 삽입할 수 있는 데이터의 양은 LSB 방식에 비해 적지만, LSB 방식에서 나타날 수 있는 이미지 품질의 왜곡을 최소화한다는 점에서 DCT 기반 스테가노그래피 방식이 권장된다[16]. 그러나 DCT는 이미지 전체가 아닌 블록단위로 적용되므로 그래픽 결함이 발생할 수 있다[15].

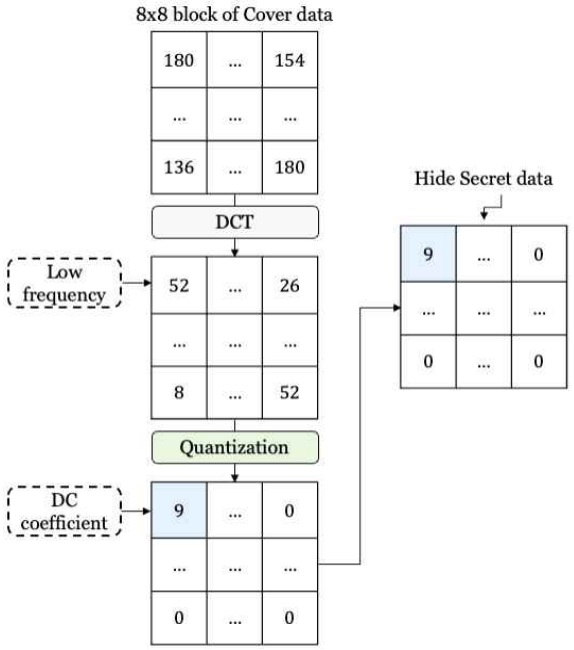

그림 3 DCT 방식

DCT는 인간의 시각 인지 한계를 이용해 주파수 성분을 변환하는 방식으로 동작한다[17,18]. DCT 기반의 스테가노그래피 임베딩 과정은 그림 3과 같다. 우선 커버 이미지를 8x8의 픽셀 블록으로 분할한다. 다음은 왼쪽에서 오른쪽으로, 위에서 아래로 가면서 각 픽셀 블록에 DCT를 수행한 후, 양자화 테이블을 통해 압축한다. 결과적으로 각 픽셀 블록들은 고주파와 저주파 성분으로 나뉘게 된다. 이때, 이미지의 대부분을 차지하는 저주파 성분 (Discrete Cosine coefficient)에 많은 정보가 몰리게 되는데, 저주파 성분에 정보를 숨기게 된다면 화질 저하가 오게 되어 안정적이지 않다. 따라서 고주파 성분인 AC 영역 (DC를 제외한 영역)에 데이터를 숨김으로써 왜곡이 적은 스테가노그래피가 가능해진다. 대표적인 예시로 각 블록의 DC 계수를 계산한 뒤, 나머지 LSB들을 비밀 메시지로 대체함으로써 이미지에 데이터를 숨기는 방식이다[16,19]. 이와 같이 DCT 기반의 스테가노그래피는 육안으로 구분하기 어려운 수준에서 데이터를 일부만 변경한다[14,16]. 위에서 살펴본 바와 같이 스테가노그래피를 위해 다양한 임베딩 방식들이 사용되고 있고, 표 2와 같이 커버 오브젝트의 타입에 따른 스테가노그래피 툴들이 오픈소스 형태로 공개되어 있어서 어렵지 않게 활용할 수 있다.

표 2. 스테가노그래피 오픈소스

Method	Tool	Cover Object Type
Steganography	Openstego	Image
	SteganPEG	Image
	Steghide	Image/Audio

2. 인공지능 기반의 스테가노그래피

기존 스테가노그래피 기술들의 품질을 보완하기 위해 인공지능 기반의 스테가노그래피 기술들이 제안되기 시작하였다. 해당 방식들은 데이터의 품질은 향상시켰으나 높은 계산 복잡도나 많은 데이터를 숨길 수 없다는 한계점을 극복하지 못하였다. [20]를 시작으로 컨볼루션 신경망 (Convolutional Neural Network, CNN) [21] 기반의 기술들이 제안되었다. 이후에는 더 나은 성능을 위해 [22]을 시작으로 생성형 적대적 신경망 (Generative Adversarial Network, GAN) [23]을 활용한 스테가노그래피 기술들이 연구되었다.

가. CNN 기반

이미지 처리에 적합한 신경망인 CNN을 활용한 스테가노그래피 기술들은 그림 4와 같이 주로 CNN 기반의 인코더와 디코더가 결합된 구조를 사용한다. 스테가노그래피를 위한 공통적인 동작 과정은 다음과 같다. 커버 오브젝트와 비밀 데이터를 병합한 후 인코

더 네트워크에 입력하면 커버 오브젝트에 비밀 데이터를 임베딩한 형태의 스테고 오브젝트가 만들어진다. 그리고 생성된 스테고 데이터를 다시 디코더 네트워크에 입력하면 커버 데이터에 임베딩 했던 비밀 데이터가 복원된다. 인코더와 디코더 네트워크는 독립적으로 학습되는 것이 아니라 인코더의 입력부터 디코더의 출력까지 end-to-end 방식으로 학습된다.

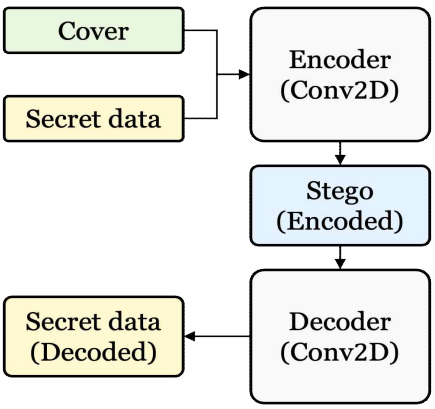

그림 4 딥러닝 기반의 스테가노그래피를 위한 CNN 구조 [24]

[24]에서는 은닉하고자 하는 비밀 데이터가 이미지인 경우에 대한 스테가노그래피 기술을 제안하였다. 즉, 적은 용량의 텍스트 데이터를 숨기는 것이 아니라 동일한 크기의 이미지를 숨기는 것이기 때문에 임베딩 용량이 높은 편에 속한다. 제안 기술은 그림 4의 구조를 기반으로 하였으며, 각 모델의 구조는 Highway[25], ResNet[26], Inception[27]에서 영향을 받은 컨볼루션 레이어 기반의 잔여 네트워크 블록을 사용하였다. 해당 기술의 왜곡 및 이미지 품질을 평가한 결과는 다음과 같다. 커버 이미지와 스테고 이미지의 히스토그램을 비교한 결과, CNN 기반의 스테가노그래피 방식이 3-bit LSB 방식보다 더 유사한 히스토그램 (더 적은 왜곡)을 가졌다. 따라서 기존 LSB 방식에 비해 더 강력한 스테고 오브젝트를 생성하였음을 알 수 있다. 그러나 이미지에서 텍스처가 풍부하지 않은 부분 (검정색, 흰색) 에서는 약간의 노이즈가 발생하였다. 저자들은 이러한 한계점은 VAE (Variational Auto Encoder) [28] 또는 GAN과 같은 정교한 생성 능력을 가진 신경망을 통해 극복할 수 있을 것으로 예상하였다.

[29]는 [20]의 개선된 버전이며, 이미지에 이미지를 숨기는 방식을 제안하였다. 또한, 한 개가 아니라 여러 개의 이미지를 하나의 커버 이미지에 은닉하는 경우에 대한 실험도 진행하였다. 해당 연구에서도 그림 4의 구조를 활용하였고 인코더와 디코더 네트워크 이전에 입력 데이터 전처리를 위한 네트워크를 추가하였다. 전처리 네트워크에서는 숨겨야 할 비밀 데이터의 RGB 채널 (3채널)에 대한 변환을 수행하여 이미지를 7개의 채널으로 변경한다. 이후, 변환된 비밀 데이터와 커버 오브젝트를 결합하여 인코더에 입력한다. 인코더는 커버 오브젝트에 비밀 데이터를 임베딩하여 스테고 오브젝트가 생성한다. 생성된 스테고 오브젝트는 디코더에 입력되고, 디코더는 스테고 오브젝트에서 비밀 데이터를 추출해낸다. 그리고 이러한 전체 과정은 end-to-end 방식으로 하나의 네트워크처럼 학습된다. 일반적인 임베딩 방식과 더불어 비밀 데이터를 난독화하여 숨긴 뒤 복구하는 방식도 제안하였다. 이때, 난독화를 위해서는 이미지 픽셀을 전치 (Permutation)하는 방식을 선택하였다. 난독화된 데이터를 은닉하는 경우, 스테고 이미지의 색감이 약간 달라지긴 했지만 육안으로 보기에 노이즈가 감지되지 않음을 보였다. 제안 기술에 대한 임베딩 성능을 평가한 결과, 비밀 데이터가 스테고 이미지의 모든 채널에 고르게 퍼져서 은닉되었음을 확인하였다. 즉, LSB 분석과 같은 간단한 스테그아날리시스로는 비밀 데이터를 복구할 수 없도록 임베딩 하였다. 이러한 임베딩을 위해 비밀 데이터를 확산 시키는 능력은 CNN 필터의 크기 및 구조에 따라 달라짐을 밝혔다. 또한, 저자들은 스테그아날리시스에 대한 저항성을 살펴보기 위해 딥러닝 분류 모델을 통해 비밀 데이터의 존재 여부를 탐지하였다. 그 결과, LSB 임베딩 보다 5~9% 더 낮은 정확도를 보이며 강력한 임베딩 능력을 가짐을 확인하였다.

나. GAN 기반

GAN 기반의 스테가노그래피는 GAN의 적대적 학습 구조를 스테가노그래피와 스테그아날리시스의 적대적 관계에 적용하며, 커버 데이터를 생성하는 방식과 스테고 데이터를 생성하는 방식으로 나뉜다.

(1) 스테가노그래피를 위한 커버 오브젝트를 생성하는 방식

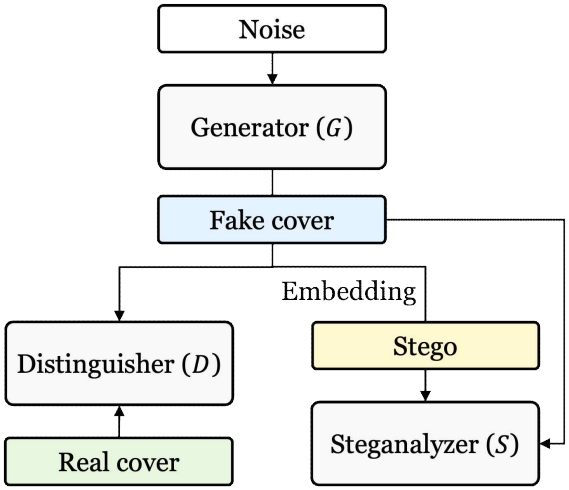

그림 5 커버 오브젝트를 생성하는 GAN 기반의 스테가노그래피 [22]

첫 번째 관점은 스테그아날리시스에 덜 민감한 커버 오브젝트를 만드는 것에 초점을 둔다. [22]은 GAN 구조를 스테가노그래피에 적용한 초기 연구이며, 그림 5와 같이 구성된다. 전체 네트워크는 3개의 신경망 모델 (G, D, S)로 구성된다. G는 랜덤 벡터를 입력으로 받아 실제 커버 이미지와 비슷한 가짜 커버를 생성한다. D는 G가 생성한 가짜 커버 데이터와 실제 커버 데이터를 입력받은 후, 가짜 커버 이미지와 진짜 커버 이미지를 구분하는 역할을 한다. S는 G가 생성한 가짜 커버 이미지에 비밀 데이터를 임베딩 (LSB와 같은 기존의 임베딩 알고리즘)하여 생성된 스테고 이미지와 가짜 커버 이미지를 입력 받은 후, 스테고 이미지와 가짜 커버 이미지에 대한 분류를 수행한다. 손실 함수로는 기본적인 GAN의 손실함수 구조를 사용하며, G와 D의 손실과 G와 S의 손실을 더하는 구조이다. 따라서 G와 D간의 학습을 통해 실제 커버 이미지와 유사한 이미지를 생성할 수 있게 되고, G와 S간의 학습을 통해 G는 스테가노그래피에 적절한 커버 데이터를 생성하게 된다. 이러한 방식으로 스테가노그래피에 적합한 커버 이미지를 생성할 수 있으며, 해당 연구를 기반으로 성능을 개선시킨 사례[30]도 존재한다.

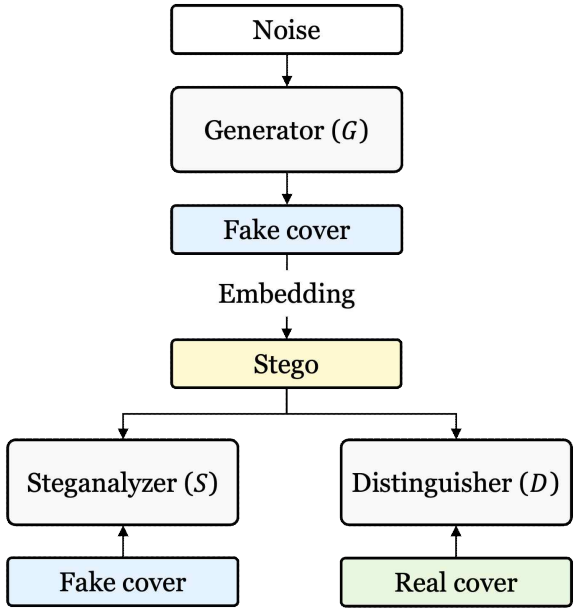

그림 6 커버 오브젝트를 생성하는 GAN 기반의 스테가노그래피 [31]

[31]은 [22,30]과는 조금 다른 관점에서의 커버 데이터 생성 기술을 제안하였다. 그림 6은 제안 기술의 전체 구조를 보여주며, [22]과 동일하게 G, D, S로 구성되어있다. G가 커버 데이터를 생성하는 것은 동일하지만, D의 기능에서 차이점이 존재한다. 해당 연구에서는 G가 생성한 가짜 커버 데이터에 비밀 데이터를 임베딩하여 스테고 이미지를 만든 후, 해당 이미지를 D와 S에 입력한다. D는 입력 받은 스테고 이미지와 실제 커버를 구별하는 역할을 하고, S는 스테고 이미지와 가짜 커버 이미지를 구별하는 기능을 수행한다. 이를 통해 G는 자신이 생성한 가짜 커버 데이터를 기반으로 스테가노그래피를 수행하는 경우에 대해 다음과 같이 2가지 성질을 만족하게 된다. 첫 번째는 육안으로 구별이 되지 않는 커버를 생성하는 것이다. 두 번째는 스테그아날리시스에 강력한 스테가노그래피가 가능해지는 커버를 생성하는 것이다. 다시 말하면, G가 생성한 가짜 커버는 실제 커버와 유사하지 않을 수도 있으며, 스테가노그래피가 적용되었을 때 비로소 실제 커버와 유사해지는 것이다. 가짜 커버 데이터를 생성하는 것은 앞서 설명한 기법[22,30]과 동일하지만 D의 기능을 변형하는 방식으로도 스테가노그래피에 적합한 커버 데이터를 생성할 수 있다.

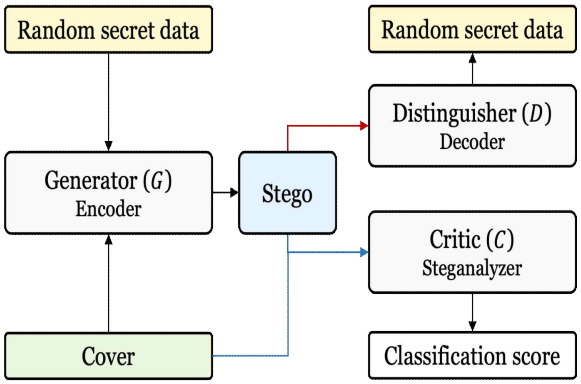

그림 7 스테고 오브젝트를 생성하는 GAN 기반의 스테가노그래피 [32]

(2) 스테고 오브젝트를 생성하는 방식

GAN 기반의 스테가노그래피에 대한 두 번째 관점은 스테고 오브젝트를 생성하는 것에 초점을 둔다. [32]은 이러한 관점에서의 GAN 기반의 스테가노그래피 방식을 제안하였다.

그림 7은 해당 논문에서 제안된 기법의 전체 구조를 보여준다. 총 3개의 신경망 모델 (G, D, C)로 구성된다. G는 스테고 이미지를 생성하는 인코더 역할을 수행한다. D는 CNN 기반의 스테가노그래피와 같이 스테고 이미지로부터 숨겨진 비밀 데이터를 복원하는 디코더 역할을 한다. 마지막으로 C는 스테고 이미지와 실제 커버 이미지를 입력 받아 분류하는 스테그아날리시스 기능을 수행하며 분류 성능을 출력한다. 기본적인 GAN의 구조에서는 G가 랜덤 벡터를 시드로써 사용하지만 제안 기법에서는 커버 데이터와 비밀 데이터를 입력받아야 한다. 이를 위해 해당 모델에서는 GAN의 seed인 랜덤 벡터를 은닉해야할 비밀 데이터로 사용하였다. 그리고 CNN 기반의 스테가노그래피와 유사하게 커버 이미지와 비밀 데이터를 함께 G에 입력한다. 입력 받은 두 데이터를 기반으로 G는 스테고 이미지를 생성한다. 즉, G-D, G-C 간의 학습을 통해 이미지 왜곡이 적고 스테그아날리시스에도 강력한 스테고 이미지를 생성하게 된다.

해당 방식은 커버데이터를 생성하는 방식과는 다르지만, 두 방법 모두 스테가노그래피의 성능을 향상시키기 위해 사용될 수 있다.

Ⅲ. 분 석 및 논 의

본 장에서는 앞서 살펴본 여러 스테가노그래피 기술에 대한 전체적인 장점과 제약 사항 및 미래 연구 방향 등을 살펴본다.

기존에는 커버 데이터의 통계적 특성을 분석하여 시각적으로 유사한 스테고 데이터를 생성해내는 기술들이 연구되어왔다. 그러나 최근에는 인공지능 기술이 발달함에 따라 기존 스테가노그래피의 임베딩 용량의 증가와 품질 개선을 위해 딥러닝 기반의 스테가노그래피 기술들이 제안되기 시작하였다. 딥러닝은 고차원 데이터에 대한 복잡한 패턴과 특징을 인식하고 추출할 수 있으므로 커버 오브젝트의 특성을 파악하고 분석한 내용을 기반으로 데이터를 더욱 정교하게 삽입하고 분석할 수 있다. 또한, 딥러닝 기술은 이미지나 비디오와 같이 다양한 타입의 데이터를 처리할 수 있으므로 여러 종류의 커버 데이터에 적용할 수 있다. 이러한 특성으로 인해 딥러닝 기반의 스테가노그래피는 기존 스테가노그래피 기술의 한계점을 극복할 수 있다.

딥러닝 기반의 스테가노그래피 모델들은 초기에는 CNN 기반의 오토인코더를 사용하였으나 최근에는 품질 향상을 위해 GAN 모델이 사용되고 있으며, 이와 같이 점점 더 정교한 생성 능력을 갖는 딥러닝 기술들이 적용되고 있음을 알 수 있다. 실제로 딥러닝이 적용된 스테가노그래피 기술들을 살펴본 결과, 데이터의 왜곡이나 이미지 품질을 고려할 때 많은 성능 향상이 있었다. 그러나 딥러닝 기술의 특성 상 이러한 작업을 수행하기 위해서는 계산 복잡도 및 메모리 사용량이 매우 크다는 단점이 존재한다.

특히 생성형 네트워크의 경우 일반적인 분류 작업보다 정교하고 복잡하므로 많은 파라미터가 필요해진다. 따라서 앞으로는 임베딩 성능을 유지하면서 네트워크의 파라미터 및 계산 복잡도를 감소시킬 수 있는 딥러닝 기술들이 추후 스테가노그래피 생성 기술에 적용되어야 할 것으로 생각된다.

Ⅲ. 결 론

본 논문에서는 스테가노그래피에 관한 기술 동향과 각 기술들의 장단점 및 제약 사항 등에 대해 살펴보았다. 최근 딥러닝 기술이 발전하면서 스테가노그래피 성능 개선을 위해 활용되고 있는 추세이다. 딥러닝 기술을 사용하면 더욱 고차원 데이터에 대한 분석 및 학습이 가능하게 되어 더욱 정교한 스테고 오브젝트를 생성할 수 있게 된다. 따라서 딥러닝 기반의 스테가노그래피 기술을 사용하면 기존 스테가노그래피 기술의 한계점인 임베딩 용량과 이미지 품질을 개선할 수 있게 되었다. 그러나, 데이터를 생성해내는 딥러닝 네트워크이기 때문에 많은 메모리 및 연산 복잡도가 요구된다. 따라서 향후에는 임베딩 용량과 성능을 유지하는 한에서 더욱 경량화 된 스테가노그래피 모델을 생성하기 위한 연구들이 필요할 것으로 생각된다.

REFERENCES

[1] DJEBBAR, Fatiha, “Securing IoT Data Using Steganography: A Practical Implementation Approach,” *Electronics,* 10.21: 2707, 2021.

[2] KHARI, Manju, et al., “Securing data in Internet of Things (IoT) using cryptography and steganography techniques,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 50, Issue 1, PP. 73-80, Jan. 2020.

[3] Akhtar, Nadeem, Pragati Johri, and Shahbaaz Khan. "Enhancing the security and quality of LSB based image steganography," *2013 5th International Conference and Computational Intelligence and Communication Networks,* Mathura, India, Sep. 2013.

[4] Jian, Chua Teck, et al., "Audio steganography with embedded text," *IOP Conference Series: Materials Science and Engineering,* Vol. 226. No. 1, Aug. 2017.

[5] Kim, Jaeyoung, et al., "A statistical approach for improving the embedding capacity of block matching based image steganography," *Journal of Broadcast Engineering,* Vol. 22, Issue 5, pp. 643-651, 2017.

[6] Kavitha, Kavita Kadam, Ashwini Koshti, and Priya Dunghav, "Steganography using least significant bit algorithm," *International Journal of Engineering Research and Applications (IJERA)*, Vol. 2, Issue 3, pp. 2248-9622, May-Jun 2012.

[7] Gupta, Shailender, Ankur Goyal, and Bharat Bhushan, "Information hiding using least significant bit steganography and cryptography," *International Journal of Modern Education and Computer Science,* 4.6: 27, Jun. 2012.

[8] Wu, Da-Chun, and Wen-Hsiang Tsai, "A steganographic method for images by pixel-value differencing," *Pattern recognition letters,* Vol. 24, Issues 9 - 10, pp. 1613-1626, Jun. 2003.

[9] Nurdiyanto, Heri, et al., "Enhanced pixel value differencing steganography with government standard algorithm," *2017 3rd International Conference on Science in Information Technology (ICSITech),* Bandung, Indonesia, Oct. 2017.

[10] H.-C. Wu, et al., “Image steganographic scheme based on pixel-value differencing and LSB replacement methods,” *IEE Proceedings-Vision, Image and Signal Processing*, Vol. 152, Issues 5, pp. 611-615, Oct. 2005.

[11] C.-H. Yang, et al., “Capacity-raising steganography using multi-pixel differencing and pixel-value shifting operations,” *Fundamenta Informaticae*, vol. 98, no. 2-3, pp. 321- 336, 2010.

[12] V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” *2012 IEEE International Workshop on Information Forensics and Security(WIFS)*, Vol.2, pp. 234 - 239, 2012.

[13] Holub, V., Fridrich, J., “Digital image steganography using universal distortion,” *Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security*, pp. 59 - 68, Montpellier, France, Jun. 2013.

[14] Nam, Soo-Tai and Chan-Yong Jin, "Rebuilding of Image Compression Algorithm Based on the DCT (discrete cosine transform)," *Journal of the Korea Institute of Information and Communication Engineering,* vol. 23, no. 1, pp. 84-89, 2019.

[15] Sahar A. El_Rahman, “A comparative analysis of image steganography based on DCT algorithm and steganography tool to hide nuclear reactors confidential information,” *Computers & Electrical Engineering,* Vol. 70, pp. 380-399, Aug. 2018.

[16] Dr. Ekta Walia, Payal Jain, Navdeep, “An Analysis of LSB & DCT based Steganography,” *Global Journal of Computer Science and Technology*, Vol. 10, Issue 1, pp.4-8, Apr. 2010.

[17] A.M.Raid, W.M.Khedr, M. A. El-dosuky and Wesam Ahmed, “Jpeg Image Compression Using Discrete Cosine Transform - A Survey,” *International Journal of Computer Science & Engineering Survey(IJCSES)*, Vol. 5, No. 2, Apr. 2014.
[18] R. Reininger and J. Gibson, “Distributions of the Two-Dimensional DCT Coefficients for Images,” *IEEE Transactions on Communications*, Vol. 31, Issue 6, pp. 835-839, Jun. 1983.
[19] Ken Cabeen, Peter Gent, “Image Compression and the Discrete Cosine Transform,”
[20] Baluja, Shumeet, "Hiding images in plain sight: Deep steganography," *Advances in neural information processing systems 30*, 2017.
[21] Albawi, Saad, Tareq Abed Mohammed and Saad Al-Zawi, "Understanding of a convolutional neural network," *2017 international conference on engineering and technology (ICET)*. pp. 1-6, Antalya, Turkey, Aug. 2017.
[22] Volkhonskiy, Denis, Ivan Nazarov and Evgeny Burnaev, "Steganographic generative adversarial networks," *Twelfth international conference on machine vision (ICMV 2019),* Vol. 11433, pp. 991-1005, Jan. 2020.
[23] Goodfellow, Ian, et al., "Generative adversarial networks," *Communications of the ACM,* Vol. 63, no. 11, pp. 139-144, Oct. 2020.
[24] Wu, Pin, Yang Yang and Xiaoqiang Li, "Stegnet: Mega image steganography capacity with deep convolutional network," *Future Internet,* Vol. 10, Issue 6, pp. 54, Jun. 2018.
[25] Srivastava, et al., "Highway networks," arXiv:1505.00387, Nov. 2015.
[26] He, Kaiming, et al., "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, Las Vegas, NV, USA, Jun. 2016.
[27] Szegedy, Christian, et al., “Rethinking the inception architecture for computer vision,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818-2826, Las Vegas, NV, USA, Jun. 2016.
[28] Kingma, Diederik P. and Max Welling, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning,* Vol. 12, No. 4, pp. 307-392, Nov. 2019.
[29] Baluja, Shumeet, "Hiding images within images," *IEEE transactions on pattern analysis and machine intelligence,* Vol. 42, Issue 7, pp. 1685-1697, Feb. 2019.
[30] Shi, Haichao, et al., "SSGAN: Secure steganography based on generative adversarial networks," *Advances in Multimedia Information Processing - PCM 2017,* pp. 534-544, Harbin, China, September 28-29, 2017.
[31] WANG, Yaojie, Ke NIU and Xiaoyuan YANG, "Information hiding scheme based on generative adversarial network," *Journal of Computer Applications,* Vol. 38, Issue 10, pp. 2923-2928, 2018.
[32] Zhang, Kevin Alex, et al., "SteganoGAN: High capacity image steganography with GANs," arXiv:1901.03892, 2019.

저 자 소 개

김 현 지(학생회원)

2020년 2월: 한성대학교 IT 응용시스템 공학과 학사 졸업.
2022년 2월: 한성대학교 IT융합공학과 석사 졸업.
2023년 3월: 한성대학교 정보컴퓨터공학과 박사과정.
<주관심분야 : 정보보안, 인공지능, 양자 컴퓨팅>

임 세 진(학생회원)

2022년 2월: 한성대학교 컴퓨터공학과 학사 졸업.
2022년 3월: 한성대학교 IT융합공학과 석사과정.
<주관심분야 : 양자 컴퓨터, 인공지능 보안, 정보보안>

김 덕 영(학생회원)

2019년 2월: 한성대학교 디자인아트평생교육원 인테리어디자인과 학사 졸업.
2023년 3월: 한성대학교 융합보안학과 석사과정.
<주관심분야 : 정보보안, 인공지능>

윤 세 영(학생회원)

2020년 3월: 한성대학교 IT 융합공학부 학사과정.
<주관심분야 : 인공지능, 디지털 포렌식>

서 화 정(종신회원)

2010년 2월: 부산대학교 컴퓨터공학과 학사 졸업.
2012년 2월: 부산대학교 컴퓨터공학과 석사 졸업.
2016년 2월: 부산대학교 컴퓨터공학과 박사 졸업.
2017년 3월: 싱가포르 과학기술청
2023년 2월: 한성대학교 IT융합공학부 조교수
2023년 3월: 한성대학교 융합보안학과 부교수
<주관심분야 : 암호구현, 정보보안>