creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

[UCI]I804:21031-200000367607

공 학 석 사 학 위 논 문

복수의 스테가노그래피 알고리즘 분류를 위한 계층적 CNN구조 기반의 스테그아날리시스

2021년 2월

부 경 대 학 교 대 학 원

인 공 지 능 융 합 학 과

강 상 훈

공 학 석 사 학 위 논 문

복수의 스테가노그래피 알고리즘 분류를 위한 계층적 CNN구조 기반의 스테그아날리시스

지도교수 박 한 훈

이 논문을 공학석사 학위논문으로 제출함.

2021년 2월

부 경 대 학 교 대 학 원

인 공 지 능 융 합 학 과

강 상 훈

강상훈의 공학석사 학위논문을 인준함.

2021년 2월 19일

위 원 장 공학박사 문 광 석 (인)

위 원 공학박사 정 완 영 (인)

위 원 공학박사 박 한 훈 (인)

목 차

그 림 목 차 ····· iii
표 목 차 ····· v

Ⅰ 서 론 ····· 1

Ⅱ 관련 이론 및 연구 ····· 6
2.1 영상 스테가노그래피 ····· 6
2.1.1 LSB 스테가노그래피 ····· 7
2.1.2 PVD 스테가노그래피 ····· 8
2.1.3 WOW 스테가노그래피 ····· 11
2.1.4 UNIWARD 스테가노그래피 ····· 13
2.1.5 MIPOD 스테가노그래피 ····· 15
2.1.6 스테가노그래피 알고리즘의 특성 ····· 15
2.2 스테그아날리시스 ····· 18
2.2.1 수제특징 기반의 스테그아날리시스 ····· 19
2.2.2 CNN 구조 기반의 스테그아날리시스 ····· 21
2.3 CNN 기반의 스테그아날리시스 네트워크 구성 ····· 23
2.3.1 CNN 구조 ····· 23
2.3.2 CNN구조의 깊이와 성능의 상관관계 ····· 26
2.3.3 ResNet 구조 기반의 방법들 ····· 32
2.3.4 DenseNet 구조 ····· 36
2.3.5 PeleeNet 구조 ····· 38
2.4 다수의 스테가노그래피 알고리즘 분류 ····· 40

Ⅲ 계층적 CNN 구조의 스테그아날리시스 ····· 43
3.1 하이퍼 파라미터 ····· 44

3.2 계층적 CNN 네트워크 구성 ····· 46

Ⅳ 실험 방법 및 결과 ····· 49
4.1 동시 분류 ····· 49
4.2 계층 분류 ····· 51
4.2.1 4계층 분류 ····· 52
4.2.2 3계층 분류 ····· 54
4.2.3 계층 분류 결과 ····· 55

Ⅴ 결 론 ····· 57

Ⅵ 향후 계획 ····· 59

참 고 문 헌 ····· 61

PUKYONG NATIONAL UNIVERSITY 부경대학교

그 림 목 차

그림 2.1 데이터 삽입 전후의 영상과 차영상 ····· 7
그림 2.2 LSB의 데이터 삽입 과정 ····· 8
그림 2.3 PVD의 데이터 삽입 과정 ····· 10
그림 2.4 WOW에서 사용한 방향성 고역통과 필터 ····· 12
그림 2.5 미러 패딩 ····· 13
그림 2.6 스테가노그래피의 적용에 따른 데이터 분포 변화 비교 14
그림 2.7 고역통과 필터의 적용 ····· 19
그림 2.8 선형, 비선형의 30개의 SRM 고역통과 필터 ····· 21
그림 2.9 XuNet의 구조와 XuNet에서 사용한 고역통과 필터 ···· 22
그림 2.10 합성 곱 연산의 적용 ····· 24
그림 2.11 최댓값 풀링과 평균값 풀링의 비교 ····· 25
그림 2.12 대표적인 활성화 함수 ····· 26
그림 2.13 Residual 구조 ····· 30
그림 2.14 VGG구조 기반의 Plain구조와 Residual구조 ····· 30
그림 2.15 [20]에서 사용한 Residual 구조 ····· 31
그림 2.16 [20]의 네트워크 구조 ····· 32
그림 2.17 변형 Residual 구조 ····· 33
그림 2.18 [36]의 변형 Residual 구조 ····· 34
그림 2.19 [37]의 Residual 구조 ····· 34
그림 2.20 [38]의 Residual 구조 ····· 35
그림 2.21 Dense Block의 구조 ····· 37
그림 2.22 DenseNet 기반의 네트워크 ····· 38
그림 2.23 DenseNet과 PeleeNet의 Dense블록 ····· 38
그림 2.24 [41]의 Stem Blcok ····· 39
그림 2.25 [41]기반의 네크워크 ····· 40

그림 3.1 단일 네트워크와 제안방법의 계층적 네트워크 구조 …· 47

PUKYONG NATIONAL UNIVERSITY
부경대학교

표 목차

표 2.1 PVD에서 사용한 레인지 테이블 ····· 9
표 2.2 PeleeNet으로 학습한 구조들의 교차검증 ····· 16
표 2.3 VGG에서 사용한 네트워크의 구조들 ····· 28
표 2.4 WOW와 UNIWARD 교차 검증 ····· 41
표 3.1 제안 방법에서 사용한 하이퍼파리미터 ····· 44
표 4.1 단일 네트워크 기반 5진 분류 ····· 50
표 4.2 4계층 PVD 검출 ····· 52
표 4.3 4계층 LSB 검출 ····· 53
표 4.4 4계층 UNIWARD 검출 ····· 53
표 4.5 4계층 COVER, WOW, MIPOD 분류 ····· 54
표 4.6 3계층 COVER, WOW, UNIWARD, MIPOD 분류 ····· 54
표 4.7 계층적 CNN의 실험 결과 ····· 56

Hierarchical Convolution Neural Network Structure for Classifying Multiple Steganographic Algorithms

Sanghoon Kang

Department of Electronic Engineering, The Graduate School,
Pukyong National University

Abstract

Steganography is a technique that embeds data into multimedia such as photography, text, video. Among them, the method of embedding data into images is called Image Steganography. Image Steganography embeds data by manipulating pixel's value of image. Early stage of image steganography concentrated on not being founded in human's eye and capacity of data, but that kind of methods leaved statistical traces, which can easily be detected. Later, steganographic technique has been progressed much elaborately and deftly, the steganographic images become more hard to detect.

The main purpose of steganography is to deliver data without detection to third parties. By embedding secret data into innocent image which called 'cover image', steganographer makes images which called 'stego image' that looks innocent.

Since Steganography can be exploited for secret information transmission, detection of Steganography is a very important problem. So Steganography and the method for detecting Steganography has been studied steadily.

The method for detecting Steganography is called Steganalysis. Along the

growth of digital communication technology, Steganography and Steganalysis has been growed in competition. Steganalysis detects stego images by using differences between cover image and stego image. because of the tininess of steganographic traces, it's relatively hard to detect stego image. In the early stages of steganalysis, stego image could be dectect by statistical differences. but along the progresses of digital technology, steganographic method became more sophisticated and had lesser traces. For detecting such sophisticate steganography, steganalysis methods use variety tools, like high pass filter, channel knowledge, machine learning and deep learning.

These days, convolution neural network(CNN) based deep learning technique show tremendous results in many computer vision's area. Steganalysis method also adopt CNN structure for classifying cover image and stego image. In CNN based method, steganalysis shows high accuracy nearly 95%, and it's almost saturated. Despite of the high accuracy, today's steganalysis can only detect images with steganography. We can find images, but we can't find data.

In this paper, as a first step toward extracting and restoring concealed data with steganography, we aim to detect the stego image and at the same time, classify which steganography algorithm is used. For this we propose CNN-based hierarchical Steganalysis structure, which has multiple networks for classifying N-stego algorithm. In general, noise in image by Steagnography is very tiny and the difference between Steganography has high similarity. The high similarity makes classification of Steganography algorithms harder. To improve the limits of a single CNN structure in classification of multiple Steganography algorithms problem, the proposed method classifies six classes of COVER, LSB, PVD, WOW, UNIWARD and MIPOD by combining existing CNN structures into hierarchical structures.

In this paper, multiple steganography algorithm classification results are shown through optimalizing CNN structure and hierarchical structure for steganalysis. The results of the experiment show the classification results for six classes through the data in Bossbase 1.01 of the hierarchical CNN structure and Our result shows the improvement of classifying accuracy through hierarchical CNN structure.

Ⅰ.서론

스테가노그래피(Steganography)는 텍스트, 사진, 비디오와 같은 멀티미디에 매체에 전달하고자 하는 정보를 삽입하는 기법이다. 스테가노그래피는 고대 그리스에서 전쟁 당시 비밀 통신을 위해 노예의 두피에 비밀 정보를 새기고 머리카락이 자란 후에 보내어 비밀 통신을 한 사건에서 유래한다. 스테가노그래피는 보안이 필요한 통신에서 주로 사용되나 테러나 사이버 범죄와 같이 악용될 가능성이 있다. 따라서 스테가노그래피가 적용된 매체의 검출에 대한 관심이 크게 증가하였다[1]. 스테가노그래피에서 데이터 삽입 이전의 매체를 커버영상(cover image), 삽입 이후의 매체를 스테고영상(stego image)이라고 부른다. 스테가노그래피는 정보의 송신자와 수신자 이외의 제 3자가 영상의 데이터 삽입여부를 알 수 없게 데이터를 삽입하는 것을 목적으로 한다. 따라서 스테가노그래피는 데이터의 삽입으로 인한 영상의 시각적, 통계적 변화의 최소화를 목표로 한다.

디지털 통신기술의 발전과 더불어 스테가노그래피의 데이터 은닉 방법은 더 교묘해졌다. 초기 스테가노그래피 기법들 중 대표적인 알고리즘인 LSB 기반의 스테가노그래피(Least Significant Bit Substitution Steganography) 알고리즘들[2-5]과 PVD기반의 스테가노그래피(Pixel Value Difference Steganography) 알고리즘[5-9]들의 경우 데이터를 삽입할 픽셀과 주변 픽셀의 관계를 고려하지 않고 시작 픽셀부터 순차적으로 데이터를 삽입하는 방법을 사용하기 때문에 삽입 이후의 통계적 차이가 상대적으로 크게 발생해 데이터의 삽입여부가 검출이 잘 되었다. 하지만 그 이후의 HUGO(Highly Undetectable SteGO)[10]와 같이 스테그아날리시스의 검출 특징을 역이용 하거나 WOW(Wavelet Obtained Weights)[11],

UNIWARD(UNIversal WAvelet Relative Distortion in Spatial Domain)[12], MIPOD[42]과 같이 삽입될 데이터의 픽셀과 주변 픽셀의 관계를 고려해 삽입 이후의 수치적인 왜곡정도를 최소화하는 위치에 데이터를 삽입하는 스테가노그래피 알고리즘의 경우 데이터 삽입여부의 검출이 어려워졌다.

스테그아날리시스(Steganalysis)는 스테가노그래피 기법이 적용된 영상을 검출하는 기법을 의미한다. 스테그아날리시스는 스테가노그래피의 적용으로 인한 시각적, 수치적 변화를 통해 스테가노그래피가 적용된 영상을 검출한다. LSB, PVD와 같은 초기 스테가노그래피 기법들은 데이터의 삽입으로 인한 통계적 변화가 크게 발생했다. 따라서 RS-Analysis[13]와 같은 통계적인 검출 방식은 매우 효과적으로 스테가노그래피를 검출할 수 있었다. 스테그아날리시스의 공격을 피하기 위해 정교화된 스테가노그래피 알고리즘은 스테그아날리시스의 검출 방법을 역으로 사용해 검출을 회피하거나, 데이터 삽입으로 인한 통계적 차이와 왜곡을 최소화하는 방식을 사용했다. 정교한 스테가노그래피 기법을 검출하기 위해 스테그아날리시스는 SPAM특징(Steganalysis by Subtractive Pixel Adjacency Matrix)[14], SRM(Spatial Rich Model Steganalysis)[8]과 같은 기법을 사용한다.

SPAM은 각 영상에서 각 픽셀에 대해 3x3의 작은 블록에 대한 8방향의 픽셀 값 차이를 구한 후 SVM(Support Vector Machine)[15] 알고리즘을 통해 Cover영상과 Stego영상을 분류하는 분류기를 학습했다. SVM을 사용한 머신러닝 기반의 방법들의 경우 기존의 RS-Analysis와 같은 방법에 비해 검출 정확도에 있어 개선이 되었으나 데이터 검출을 위한 복잡한 특징 생성과정과, SVM의 적용으로 인해 검출에 많은 시간이 소요되었다.

SRM[16]은 SPAM의 특징을 역이용해 데이터를 삽입한 HUGO를 검출하기 위해 30개의 선형, 비선형의 고역통과 필터(High-pass filter)를 이용

해 영상을 전 처리하는 방법을 도입했다. 고역통과 필터의 역할은 스테가노그래피의 적용으로 인한 미소한 흔적(Stego Noise)을 확대하고 데이터가 삽입되지 않는 영상의 콘텐츠 영역을 제거하기 위해 도입되었다. SRM은 30개의 고역통과 필터로부터 다양한 방향성을 가지는 특징을 추출해 스테가노그래피가 적용된 영상에 대한 검출 정확도를 향상시킬 수 있었고, SRM의 고역통과 필터 개념은 이후의 스테그아날리시스 기법들에도 사용된다.

영상 분류에 있어 CNN(Convolution Neural Network)을 사용한 딥러닝(Deep Learning) 기반의 방법을 사용한 AlexNet[17]의 ImageNet[18] 영상 분류 분류의 큰 성능 향상 이후, 많은 영상 분류 분야에서 CNN을 사용해 영상을 분류하기 시작했다. 기존의 머신러닝 기반의 방법에 비해 딥러닝 기반의 방법은 비교적 간단한 구조를 가지나 높은 성능을 보이는 장점이 있다. 딥러닝 기반의 방법은 기존 방법들과 같이 추출할 특징을 설계하고 분류하는 방식이 아닌 Conv연산을 통해 딥러닝 구조가 분류기준을 학습하고 가중치를 스스로 최적화하는 방법을 사용한다. 딥러닝 기반의 분류 방법은 구조에서 예측한 결과와 실제 정답을 비교한 결과를 다시 구조에 적용해 예측 오류를 줄여나가는 방법을 사용한다.

영상 스테그아날리시스의 경우 CNN 구조를 사용한 XuNet(Structural Design of Convolutional Neural Networks for Steganalysis)[19]을 시작으로 많은 방법들에서 CNN구조를 이용해 Cover영상과 Stego영상을 분류했다. 다른 영상분류의 분야들과 달리 CNN기반의 스테그아날리시스는 영상을 SRM에서 사용한 것과 같은 고역통과 필터를 이용해 영상을 전 처리하는 과정을 거친다. 전처리를 위한 필터의 값은 고정된 가중치 값을 사용한다. CNN 기반의 스테그아날리시스는 커버영상과 스테고영상의 2진 분류를 통해 데이터를 검출했다. XuNet을 기점으로 다수의 스테그아날리시스는

분류 정확도 향상을 위한 구조의 최적화와 기법의 사용에 치중했다. 최근의 스테그아날리시스 방법 중 하나인 Deep Residual Learning(Deep Residual Learning for Image Steganalysis)[20]은 Conv계층의 출력에 입력을 더해주는 방법을 사용해 Conv연산으로 인한 기울기의 소실을 방지하고 입력에 대해서 더 잘 학습해 WOW와 UNIWARD에 대해서 각각 95%와 94%의 정확도로 스테고영상을 검출했다.

지금까지의 CNN 기반의 스테그아날리시스는 커버영상과 스테고영상의 2진 분류 정확도를 높이는데 집중해왔다. [20]과 다른 방법론들의 분류 결과에서 나타난 것처럼 2진 분류의 정확도는 이미 상당한 수준을 보이고 있다. 지금까지의 많은 스테그아날리시스 연구는 영상에 스테가노그래피 알고리즘의 적용 여부를 검출해왔다. 하지만 커버, 스테고의 2진 분류는 영상에 어떠한 스테가노그래피가 적용되었는지 분류하지 못한다. 또한 스테가노그래피가 적용된 영상으로부터 데이터의 추출 및 복원은 상당히 요원하다. 영상으로부터 데이터를 추출, 복원하기 위해서는 먼저 영상에 어떠한 스테가노그래피 알고리즘이 적용되었는지를 분류할 수 있어야 한다.

본 논문에서는 CNN 기반의 계층적 스테그아날리시스를 이용해 스테가노그래피 기법의 적용여부를 검출하고 스테고영상의 경우 어떠한 알고리즘의 스테가노그래피가 적용되었는지 분류하는 방법을 제안한다. 이전의 CNN기반 3진 분류구조[21]에서는 단일 CNN구조를 통해 COVER, WOW, UNIWARD를 분류했다. [21]의 결과에서 스테가노그래피 알고리즘을 동시에 분류할 경우 일정 수준의 정확도를 보이나, 클래스의 수가 증가할수록 정상적인 분류가 어려운 것을 확인하였다. 클래스의 수가 증가하면 구조를 학습시키기 위한 데이터의 수도 증가해 학습에 필요한 시간과 하드웨어의 성능 필요가 증가한다. 스테그아날리시스 문제는 스테가노그래피의 적용으로 인해 발생하는 작은 차이를 분류해야하기 때문에 복수의 스테가노그래

피 알고리즘 분류는 단일 네트워크의 CNN 구조에서 클래스의 수를 증가시키는 것만으로는 분류가 어려워진다.

본 논문에서는 [21]의 결과를 개선하고, 다양한 스테가노그래피 알고리즘들을 분류하기 위해 COVER, LSB, PVD, WOW, UNIWARD, MIPOD의 6가지 클래스를 분류하는 계층적 CNN구조를 사용한다. 본 논문에서 제안하는 계층적 CNN구조는 복수의 스테가노그래피 알고리즘을 단일 CNN구조에서 분류하는 것이 아닌 스테가노그래피 알고리즘의 특징을 분석하고 복수의 네트워크를 이용해 데이터를 계층적으로 분류한다. 복수의 네트워크로 나누어 분류하는 방법은 각 네트워크에서 분류해야할 클래스의 수를 줄여 네트워크의 분류를 위한 기준이 지나치게 복잡해지는 것을 방지할 수 있다. 또한 각 네트워크마다 서로 다른 CNN 구조를 사용해 스테가노그래피 알고리즘 마다 분류에 최적화된 CNN구조를 사용해 기존의 단일구조 기반의 방법에 비해 정확도를 향상시킬 수 있다.

Ⅱ.관련 이론 및 연구

2.1 영상 스테가노그래피

스테가노그래피는 사진이나 텍스트, 비디오와 같은 멀티미디어 매체의 데이터를 변조해 전달하려는 데이터를 삽입하는 기법이다. 매체 자체를 암호화해서 전달하는 크립토그래피(Cryptography)[22]와 달리 스테가노그래피는 삽입된 데이터의 존재 여부를 은닉한다. 따라서 수신자와 송신자를 제외한 제 3자는 데이터의 삽입여부를 알아차리기 쉽지 않다. 스테가노그래피는 데이터의 삽입여부를 은닉하기 위해 삽입으로 인한 매체의 왜곡을 최소로 하는 방향으로 데이터를 삽입한다.

영상 스테가노그래피는 여러 가지 멀티미디어 매체 중 사진과 비디오 같은 영상 매체에 스테가노그래피는 영상에 데이터를 삽입하는 방법을 의미한다. 영상은 N×M의 픽셀들의 집합으로 정의되는데 영상 스테가노그래피는 이 픽셀들의 값을 변조해 데이터를 삽입한다. 그림 2.1은 스테가노그래피의 알고리즘 중 하나인 WOW를 이용해 영상에 데이터를 삽입한 것이다. 그림 2.1-(a)는 데이터가 삽입되지 않은 커버영상이고 그림 2.1-(b)는 데이터가 삽입된 스테고영상이다. 그림 2.1-(a)와 그림 2.1-(b)는 시각적으로 완전히 동일하게 보인다. 하지만 그림 1-(c)의 두 영상에 대한 차 영상을 보면 영상에 데이터의 삽입으로 인한 변화가 있는 것을 볼 수 있다.

(a) (b) (c)

그림 2.1 데이터 삽입 전후의 영상과 차영상 (a)데이터 삽입 이전의 커버 영상, (b)데이터 삽입 이후의 스테고 영상, (c)커버 영상과 스테고 영상의 차이 영상

2.1.1. LSB 스테가노그래피

LSB 스테가노그래피는 픽셀의 값을 비트 단위로 변조하여 데이터를 삽입한다. 예를 들어 픽셀이 '233'의 값을 가지고 삽입하려는 데이터가 '0'이라고 하면, '233'은 2진수로 '11101001'이 된다. 여기서 가장 하위 픽셀인 '1'을 삽입하려는 데이터인 '0'으로 대체하면 '11101000'이 되고 픽셀 값은 '232'가 된다. LSB의 데이터 삽입은 비트 단위로 이루어지기 때문에 삽입으로 인한 변화는 최대 $\pm 2^n$이 발생한다. 그림 2.2는 LSB의 데이터 삽입과정을 개략적으로 나타낸 것이다.

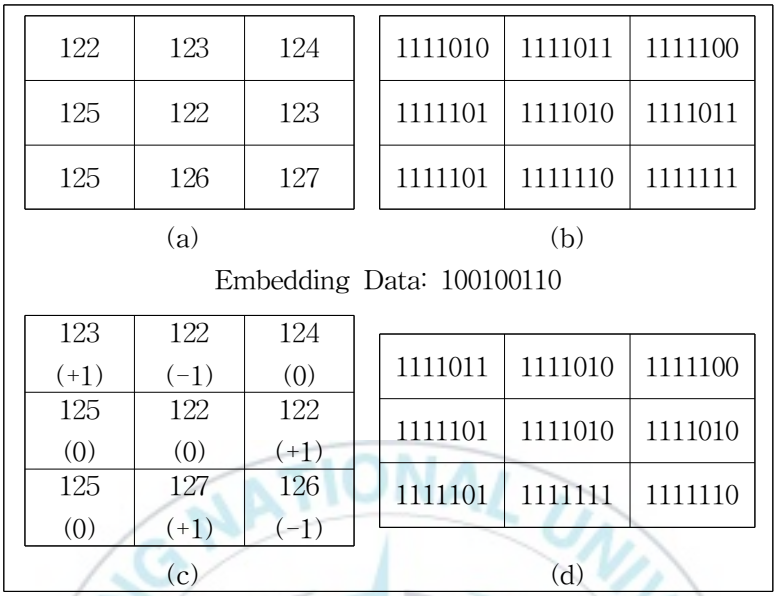


그림 2.2 LSB의 데이터 삽입 과정. (a)삽입 이전의 픽셀, (b)(a)의 2진 행렬 값, (c)삽입 이후의 픽셀, (d)(c)의 2진 행렬 값

LSB의 비트 단위로 데이터를 삽입하고 사용하는 비트의 수에 따라서 n-LSB라고 부른다. 데이터의 삽입으로 인한 영상의 왜곡을 최소화하기 위해서 LSB는 주로 1-LSB를 이용해 데이터를 삽입하는데, 이로 인해 LSB는 삽입할 수 있는 데이터의 양(Capacity)이 다른 스테가노그래피 알고리즘에 비해 상대적으로 적다.

2.1.2. PVD 스테가노그래피

PVD 스테가노그래피는 이웃하는 2개의 픽셀의 차이 값에 따라 데이터를 삽입한다. PVD는 두 픽셀의 차이 값이 커질수록 더 많은 데이터를 삽입하며, 표 2.1과 같이 차이 값의 구간을 나누어 삽입할 픽셀의 양을 결정

한다. PVD는 차이 값의 구간을 상한 값(u_k)과 하한 값(l_k)을 이용해 변화 값(m)을 인접하는 두 픽셀에 나누어 데이터를 삽입한다. PVD의 데이터 삽입은 그림2.3에 나타낸 과정을 거친다.

픽셀 값 차이(d)	l_k	u_k	삽입할 데이터의 수(n)
0 ~ 7	0	7	3
8 ~ 15	8	15	3
16 ~ 31	16	31	4
32 ~ 63	32	63	5
64 ~ 127	64	127	6
128 ~ 255	128	255	7

표 2.1 PVD에서 사용한 레인지 테이블.

PUKYONG NATIONAL UNIVERSITY 부경대학교

인접하는 두 픽셀을 $(f_i, f_{i+1}) = (35, 66)$,
삽입할 데이터를 1011010_2이라고 할 때
두 픽셀 값의 차이는,

$d = f_{i+1} - f_i = 31,$

[Table 1]의 값에 따라서, $u_k = 31, l_k = 16, n = 4$가 된다.

데이터에서 n-bit만큼 데이터를 읽어 10진수로 변환하면,
$b = 1011_2 = 11$이 된다.

$$d' = \begin{cases} l_k + b & \text{if } d \geq 0 \\ -(l_k + b) & \text{if } d < 0 \end{cases}$$ 에서 $d' = 27$

$m = d' - d$에서 $m = -4$

$$(g_i, g_{i+1}) = \begin{cases} (f_i - ceil(m/2), f_{i+1} + floor(m/2)) & \text{if } d \text{ is odd number} \\ (f_i - floor(m/2), f_{i+1} + ceil(m/2)) & \text{if } d \text{ is even number} \end{cases}$$ 에서,

삽입 이후의 픽셀 값은 (37, 64)이 된다.
삽입 이후에 차이 값이 d'과 같으므로 삽입된 데이터를 추출할 수 있다.

그림 2.3 PVD의 데이터 삽입 과정.

PVD는 두 픽셀의 차이에 비례해서 데이터를 삽입하기 때문에 LSB에 비해서 많은 데이터를 삽입할 수 있다. 하지만 PVD 방법은 데이터의 삽입으로 인한 영상의 왜곡이 매우 크다. PVD 스테가노그래피는 인접한 두 픽셀을 이용하는 방법 이외에도 3방향 픽셀을 이용한 Tri-way PVD[7], 3x3의 9개의 픽셀에서 LSB와 PVD를 혼용해 사용하는 [8,9] 등이 있다.

2.1.3. WOW 스테가노그래피

LSB와 PVD는 데이터 삽입의 목표를 삽입용량(Capacity)과 삽입 이후의 영상의 품질(PSNR[23])에 두었다. 하지만 LSB와 PVD의 데이터 삽입은 영상의 픽셀들 사이의 관계를 고려하지 않고 데이터를 순차적으로 삽입했기 때문에 영상의 통계적 수치의 변화가 발생하고 검출이 쉽게 되었다. WOW는 데이터를 삽입할 때 삽입 용량이 아닌 은닉성을 목표로 한다. WOW는 영상에 데이터가 삽입되었을 때의 왜곡의 정도를 식 (2-1)과 같이 정의한다.

$$D(X,Y)=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\rho_{ij}(X,Y_{ij})|X_{ij}-Y_{ij}| \qquad (2\text{-}1)$$

X는 커버영상, Y는 스테고영상, ρ는 픽셀의 값이 X_{ij}에서 Y_{ij}로 변할 때의 픽셀의 비용(cost)을 의미한다. WOW의 데이터 삽입은 최대 ±1로 제한하여 데이터를 삽입하기 때문에 왜곡함수($D(X,Y)$)는 ρ에 비례한 값을 가진다. WOW는 각 픽셀에 +1 혹은 -1이 삽입되었을 때의 비용을 계산해 왜곡의 정도가 최소한이 되는 픽셀에 데이터를 삽입한다. WOW는 픽셀의 변화 비용을 계산하기 위해 영상에 미러패딩(mirror-padding)(그림 2.5)을 적용한 후 방향성 필터(그림 2.4)와 Cover 영상의 합성 곱(convolution) 연산을 이용해 각 영상의 잔차(Residual)를 구하고 식 (2-2), 이를 바탕으로 각 픽셀에 대한 확률 맵을 생성한다. R은 합성 곱 연산이 적용된 영상을 의미하고, K와 *는 각각 방향성 필터와 합성 곱 연산을 의미한다.

$$R^{(k)} = K^{(k)} * X \tag{2-2}$$

$$\xi_{ij}^{(k)} = |R^{(k)}| * |R^{(k)} - R_{ij}^{(k)}|^{rot} = |R^{(k)}| * |K^{(k)}|^{rot} \tag{2-3}$$

$$\rho_{ij} = (\sum_{k=1}^{n} \frac{1}{\xi^{(k)}}) \tag{2-4}$$

식 (2-2)의 합성 곱 연산을 각 방향에 적용해 식 (2-3)의 적합도(ξ)를 구한다. ξ을 식 (2-4)과 같이 합해 각 픽셀에서 데이터가 삽입되었을 때의 비용을 구한다. WOW는 픽셀과 픽셀을 둘러싼 8방향 픽셀들과의 관계를 고려해 영상의 왜곡을 최소화하는 방향으로 데이터를 삽입하기 때문에 기존의 LSB나 PVD와 같은 스테가노그래피 알고리즘들에 비해 영상에 발생하는 왜곡이 현저하게 줄어든다.

KB	$K^{(1)} = \begin{pmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{pmatrix}$
Sobel	$K^{(1)} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$, $K^{(2)} = (K^{(1)})^T$
WDFB-H	h = Haar wavelet decomp. low-pass g = Haar wavelet decomp. high-pass
WDFB-D	h = Daubechies 8 wavelet decomp. low-pass g = Daubechies 8 wavelet decomp. high-pass $K^{(1)} = h \cdot \ g^T, K^{(2)} = g \cdot \ h^T, K^{(3)} = g \cdot \ g^T$

그림2.4 WOW에서 사용한 방향성 고역통과 필터.

100	110	120
150	200	100
100	122	119

(a)

100	100	110	120	120
100	100	110	120	120
150	150	200	100	100
100	100	122	119	119
100	100	122	119	119

(b)

그림2.5 미러 패딩. (a)원본 행렬, (b)미러 패딩 적용후의 행렬

2.1.4. UNIWARD 스테가노그래피

UNIWARD 스테가노그래피는 WOW의 삽입방식을 개선해 스테그아날리시스에 의한 검출정확도를 더 낮추고 공간도메인 뿐 아니라 JPEG도메인에서도 사용가능한 범용적인 스테가노그래피 알고리즘을 제안했다. UNIWARD는 WOW에서 사용한 필터 중 WDFB-B 필터를 사용해 영상의 잔차(residual)를 구한다. UNIWARD와 WOW의 차이는 적합도의 정의에 있다. WOW는 식 (2-4)에서 Reciprocal Holder Norm을 이용해 적합도를 정의했다. 하지만 JPEG과 같은 주파수 도메인에서는 WOW에서 정의한 적합도는 데이터가 쉽게 탐지되는 고주파 영역에 낮은 비용을 부여할 가능성이 생긴다. 이는 영상에서 탐지가 쉽게 되는 고역통과 필터 영역에 데이터를 삽입해 검출 가능성이 높아지게 된다. UNIWARD에서 정의한 적합도는 WOW의 적합도 정의의 역수를 사용했다(식 (2-5)).

$$\rho_{ij} = (\sum_{k=1}^{n} \xi^{(k)}) \qquad (2\text{-}5)$$

UNIWARD는 WOW와 같이 높은 은닉성을 지니면서도 WOW의 단점을 개선하고 공간 도메인뿐 아니라 주파수 도메인에도 적용 가능한 범용적인 스테가노그래피 알고리즘을 제안했다. UNIWARD를 이용한 스테가노그래피 알고리즘 중 공간 도메인의 알고리즘을 S-UNIWARD, JPEG 도메인을 이용한 알고리즘을 J-UNIWARD라고 한다.

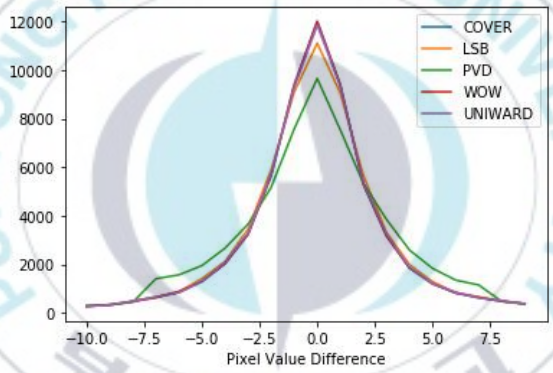


그림 2.6 스테가노그래피의 적용에 따른 데이터 분포 변화 비교.

그림 2.6은 커버영상과 각 스테가노그래피 알고리즘들이 적용된 스테고영상들의 인접 픽셀에 대한 차이의 값들의 통계이다. 앞에서 언급한 것과 같이 LSB, PVD의 분포는 COVER의 분포와 큰 차이를 보이고 WOW와 UNIWARD의 분포는 COVER와 거의 유사한 분포를 보인다. 커버영상과 분포가 유사할수록 스테고영상의 검출확률은 낮아진다. 최근의 스테가노그래피 기법들은 이와 같이 COVER의 분포와의 차이를 줄여 검출이 잘 되지 않는 것을 목표로 한다.

2.1.5. MIPOD 스테가노그래피

MIPOD은 WOW와 UNIWARD에서 정의한 왜곡함수는 각 픽셀에 데이터가 삽입되었을 때 왜곡정도를 결정하는데 이는 통계적 검출정도와 연관성이 부족한 점을 지적하고 이를 개선한 왜곡함수의 정의와 데이터 삽입 방법을 제안했다. MIPOD에서 제안한 데이터 삽입 방법은 아래의 식과 같다.

$$\begin{aligned} P(y_n = x_n + 1) &= \beta_n, \\ P(y_n = x_n - 1) &= \beta_n, \\ P(y_n = x_n) &= 1 - 2\beta_n, \end{aligned} \tag{2-6}$$

$$R(\beta) = \sum_{N}^{n=1} H(\beta_n) \tag{2-7}$$

$$H(x) = -2x\log x - (1-2x)\log(1-2x)$$

$$\beta_n = \frac{e^{-\lambda\rho_n}}{1 + 2e^{-\lambda\rho_n}} \tag{2-8}$$

식(2-6)에서 β_n은 상대적인 변화정도를 의미하고 β_n는 최대 1/3의 값을 가진다. 식(2-7)의 $R(\beta)$는 영상에 최대로 삽입 가능한 payload를 의미한다.

2.1.6 스테가노그래피 알고리즘의 특성

영상 스테가노그래피는 각 픽셀의 값을 조작해 전달하고자 하는 데이터를 영상에 삽입한다. 초기 스테가노그래피 알고리즘인 LSB와 PVD 같은 경우

데이터의 삽입 용량과 영상의 시각적 품질(PSNR)에 데이터 삽입의 주안점을 두었기 때문에 상대적으로 데이터의 검출이 쉬운 편이다. 이후 등장한 WOW와 S-UNIWARD, MIPOD과 같은 EGDE기반 스테가노그래피 알고리즘은 데이터의 삽입을 순차적이 아닌 데이터가 삽입될 위치를 결정해 데이터를 삽입하는 방법을 사용했기 때문에 초기의 스테가노그래피 알고리즘에 비해 상대적으로 검출이 어려웠다. 정교한 스테가노그래피 알고리즘일수록 영상에 발생하는 왜곡의 정도가 줄어든다.

CNN기반의 스테그아날리시스 구조에서는 구조를 학습하기 위해 다수의 커버영상과 스테고영상의 데이터를 사용하는데, 학습에 사용한 스테고영상이 어떠한 스테가노그래피 알고리즘인가에 따라 다른 스테가노그래피 알고리즘들을 잘 검출하거나 검출하지 못하는 현상이 발생한다.

본 절에서는 PeleeNet 구조를 이용해 특정한 스테가노그래피 알고리즘으로 구조를 학습하고 다른 스테가노그래피 알고리즘에 대해서 분류를 한 결과의 교차검증을 통해 스테가노그래피의 계층적 특성을 보인다. 표 3.2는 PeleeNet 구조를 이용해 각각의 스테가노그래피 알고리즘에 대해 학습 데이터와 검증 데이터를 달리 했을 때의 분류 결과이다.

Pelee	COVER	PVD	LSB	UNIWARD	WOW	MIPOD
PVD	0.9975	**0.9995**	0.0033	0.0026	0.0026	0.0026
LSB	0.9826	0.9978	**0.9486**	0.0361	0.0342	0.0265
UNIWARD	0.9352	0.2067	0.0809	**0.9212**	0.0617	0.06790
WOW	0.8208	0.9180	0.5586	0.5799	**0.7788**	0.5578
MIPOD	0.8261	0.9977	0.8554	0.5957	0.5920	**0.7026**

표2.2 PeleeNet으로 학습한 구조들의 교차검증.

표 3.2에서 각 행은 PeleeNet구조를 학습한 데이터를 의미하고, 각 열은 검증한 데이터를 의미한다. 첫 번째 행은 PVD로 학습한 구조에 COVER, PVD, LSB, UNIWARD, WOW, MIPOD의 데이터를 검출했을 때의 정확도를 의미한다. 학습 데이터와 검증 데이터가 동일할 때 각 스테가노그래피 알고리즘은 PVD, LSB, UNIWARD의 경우 약 92%이상, WOW의 경우 77%, MIPOD의 경우 약 70%정도의 분류 정확도를 보인다. PVD로 학습한 구조의 경우 COVER와 PVD의 경우 제대로 분류를 했으나 다른 스테가노그래피 알고리즘의 경우 거의 대부분 COVER로 분류했다. LSB로 학습한 구조의 경우 PVD와 LSB는 스테고영상으로 분류를 했지만 UNIWARD와 WOW, MIPOD의 경우 약 97%정도의 영상을 커버영상으로 분류했다. UNIWARD로 학습한 구조의 경우 PVD 데이터의 20%정도를 스테고영상으로 분류를 했고 나머지 알고리즘들의 경우 대부분 COVER로 분류했다. WOW는 PVD의 경우 대부분 스테고영상으로 분류를 했지만 LSB와 UNIWARD의 경우 50%정도만 스테고영상으로 분류를 했다. 이는 COVER / WOW 2진 분류의 정확도 자체가 낮아서 발생한 것으로 보인다. MIPOD의 경우도 WOW와 비슷한 경향을 보이나 UNIWARD와 WOW를 잘 분류하지 못한다.

표 3.2의 분류 정확도를 우 상단 기준으로 보면 스테가노그래피 알고리즘들의 계층적 특성을 확인할 수 있다. PVD로 학습한 구조는 LSB, UNIWARD, WOW, MIPOD을 모두 검출하지 못하고, LSB로 학습한 구조는 UNIWARD와 WOW, MIPOD을 검출하지 못한다. UNIWARD로 학습한 구조의 경우 WOW, MIPOD의 데이터를 대부분 검출하지 못하는 것을 볼 수 있다. 2진 분류 구조에서 검출하지 못한 데이터는 모두 커버영상으로 분류되기 때문에 스테가노그래피 알고리즘을 분류할 때 계층적 특징은 분류에 도움을 줄 수 있다. 예를 들어, PVD알고리즘의 경우 정확도가 거

의 100%에 가깝게 정확히 분류된다. 따라서 COVER, LSB, PVD, WOW, UNIWARD, MIPOD의 데이터를 PVD로 학습한 구조를 통해 분류를 하면 PVD는 대부분 검출되고, COVER, LSB, WOW, UNIWARD, MIPOD의 데이터만 남게 된다. 이러한 스테가노그래피의 계층적 특성은 복수의 알고리즘 분류 성능을 향상시킬 수 있다.

2.2 스테그아날리시스

스테그아날리시스는 영상에 스테가노그래피 기법이 적용되었는지 여부를 탐지하는 기법을 의미한다. 스테그아날리시스는 스테가노그래피의 적용으로 인한 미세한 통계적, 수치적 흔적을 탐지해 스테고영상을 검출한다. 스테가노그래피의 적용으로 인한 영상의 왜곡은 매우 미소하기 때문에 스테그아날리시스는 다른 문제들에 비해 상대적으로 데이터의 검출이 어렵다. 스테고영상을 검출하기 위해 초기 스테그아날라시스 방법은 영상의 픽셀 값의 통계적 차이를 이용했다. 하지만 스테가노그래피 기법이 정교해질수록 스테고영상의 분포는 커버영상의 분포와 유사하게 데이터를 삽입했고, 스테그아날리시스 방법은 이에 따라 영상에 데이터가 삽입될 확률이 높은 위치를 중심으로 분류를 수행한다. 스테그아날리시스 방법은 입력으로 들어온 영상에 고역통과 필터(High Pass Filter)를 적용에 데이터가 잘 삽입되지 않는 영상의 콘텐츠 영역을 제거하고 전경과 배경의 경계선 부분을 검출한다. 그림 2.7은 각각 영상에 고역통과 필터가 적용되기 전, 후의 영상이다. WOW, UNIWARD, MIPOD과 같은 정교한 스테가노그래피 기법은 그림 2.7-(b)와 같은 경계선영역에 데이터를 주로 삽입하기 때문에 데이터를 잘 검출하려면 경계선영역의 데이터를 바탕으로 판별을 한다. 스테그아날리시스에서 사용하는 고역통과 필터는 SRM의 30개의 선형, 비선형

의 방향성 필터를 주로 사용한다.

(a) (b)

그림 2.7 고역통과 필터의 적용. (a)원본 영상, (b)고역통과 필터가 적용된 영상

스테그아날리시스는 수제특징기반의 방법과 CNN기반의 방법으로 나누어진다. 수제특징기반 방법의 대표적인 알고리즘은 SPAM, SRM이 있고 CNN기반의 방법은 Xu와 Deep Residual Learning 등이 있다.

2.2.1. 수제특징 기반의 스테그아날리시스

수제특징 기반의 스테그아날리시스는 데이터가 삽입된 영상의 통계적 왜곡을 이용해 스테고영상을 검출한다. 그림2.6은 데이터의 삽입으로 인한 통계적 차이를 나타낸다. LSB와 같은 알고리즘의 경우 데이터의 삽입으로 최대 ±1이 변화하기 때문에 차이가 작은 영역에서의통계적 왜곡이 크게 나타난다. 수제특징 기반의 스테그아날리시스는 이러한 특성을 이용해 데이터를 검출한다. 수제특징 기반 스테그아날리시스의 대표적인 방법으로는 SPAM과 SRM이 있다.

SPAM은 수직, 수평, 대각의 8방향의 인접 픽셀 값 차이를 이용해 데이터 검출을 위한 SPAM 특징을 생성했다. SPAM 특징의 생성 과정은 각 방향에서의 인접 픽셀(식 (2-6))과의 차이를 구하고 식 (2-7) 마르코프 과정을 이용해 동시발생 행렬을 구한다. 동시 발생 행렬의 차원을 줄이기 위해 식 (2-8)의 1차 SPAM 특징과 2차 SPAM 특징으로 평균을 구한 후, SVM을 이용해 데이터를 분류한다. SPAM은 LSB와 같은 저차원의 데이터 삽입 방식을 검출하기 위해 SPAM 특징에 포함되는 인접 픽셀간의 차이를 작은 구간([-T, T])로 제한했다.

$$D_{i,j} = I_{i,j} - I_{i,j+1} \tag{2-6}$$

$$M_{u,v} = \Pr(D_{i,j+1} = u | D_{i,j} = v) \tag{2-7}$$

$$F_{1,\ldots k} = \frac{1}{4}[M^{\leftarrow} + M^{\rightarrow} + M^{\uparrow} + M^{\downarrow}]$$
$$F_{2,\ldots 2k} = \frac{1}{4}[M^{\searrow} + M^{\swarrow} + M^{\nearrow} + M^{\nwarrow}] \tag{2-8}$$

SPAM 특징은 LSB와 같이 인접 픽셀간의 변화가 발생하는 스테가노그래피 알고리즘을 잘 검출했다. HUGO는 이러한 SPAM의 특징을 역으로 이용해 SPAM에 잘 검출되지 않는 스테가노그래피 알고리즘을 사용했다.

HUGO는 SPAM 특징을 역이용하여 SPAM 특징에 검출될 수 있는 영역을 피해 데이터를 삽입한다. 이에 따라 SRM은 HUGO와 같은 알고리즘의 검출을 위해 그림 2.8의 30개의 선형, 비선형 고역통과 필터를 이용해 데이터를 검출했다. SRM은 10,000개 이상의 고차원의 특징들을 추출하고 SVM과 랜덤 포레스트(Random Forest)[24], FLD[25]를 이용한 앙상블 방법으로 데이터를 분류했다. SRM의 고역통과 필터를 통한 영상의 전처리 방법은 이후의 많은 스테그아날리시스 방법들에 사용되고 있다.

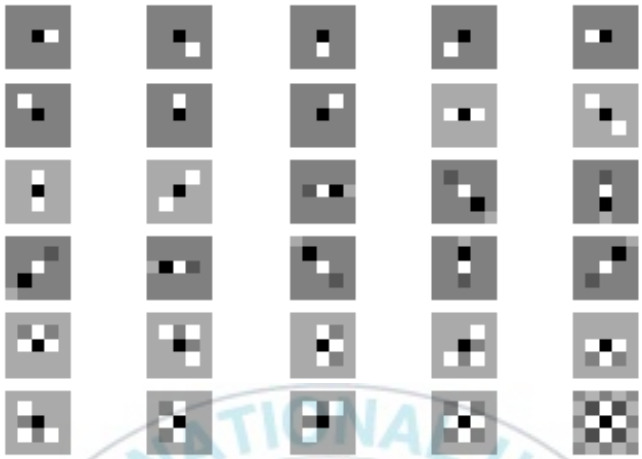


그림 2.8 선형, 비선형의 30개의 SRM 고역통과 필터.

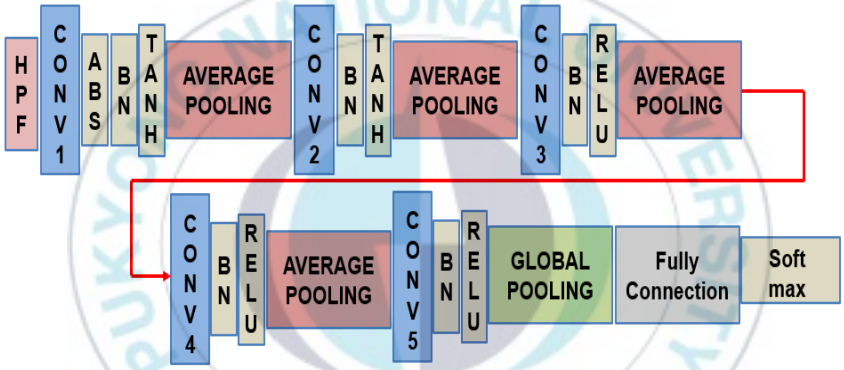
2.2.2 CNN 구조 기반의 스테그아날리시스

수제 특징기반의 스테그아날리시스는 HUGO와 같은 스테가노그래피 알고리즘을 잘 검출할 수 있었지만, 앙상블 방법을 이용한 특징 생성과정과 학습에 많은 시간이 소요되었다. AlexNet 이후 많은 스테그아날리시스 방법들이 CNN 기반의 방법을 사용한다. CNN 기반의 방법은 다수의 데이터로부터 CNN구조가 스스로 데이터를 추출하고 학습하는 과정을 통해 이전의 수제특징기반 스테그아날리시스에 비해 간결한 구조와 동시에 높은 성능을 보인다. CNN 기반의 스테그아날리시스의 대표적인 방법으로는 Xu와 Deep Residual Learning 등이 있다.

그림 2.9-(a)는 Xu의 구조를 나타낸 것이다. Xu는 512×512의 그레이스케일 이미지를 입력으로 받아 그림 2.9-(b)의 고역통과 필터로 영상을 전처리한다. 처리된 영상은 Conv계층을 거치면서 특징을 추출하고, 풀링계층을 통해 영상의 크기가 1/4로 줄어들게 된다. 입력으로 들어온 영상

(1×1×512×512)은 Conv1에서 특징 맵의 수가 8개로 증가하고, 그 외의 Conv계층에서는 2배로 증가한다. 최종 전역 풀링 계층을 적용한 후 영상은 128×1×1의 128개의 특징 맵이 되고 완전연결 계층을 거쳐 두 클래스(Cover, Stego)에 대한 2진 확률 값을 도출한다.

Xu의 특징은 첫 번째 Conv계층에서 절댓값(ABS)계층과 Tanh 활성화 함수를 사용했다. Xu는 UNIWARD에 대해 80.24%의 정확도로 기존의 SRM의 79.53%의 정확도보다 높은 성능을 보였다.



(a)

$$K_0 = \frac{1}{12}\begin{pmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{pmatrix}$$

(b)

그림 2.9 XuNet의 구조와 XuNet에서 사용한 고역통과 필터 (a)XuNet CNN구조, (b)XuNet에서 사용한 고역통과 필터

2.3 CNN 기반의 스테그아날리시스 네트워크 구성

2012년 ImageNet에서의 큰 성과를 얻은 AlexNet 이후 많은 분야에서 CNN(Convolution Neural Network)기반의 방법이 사용되었다. 기존의 영상분류 방법들이 문제해결을 위해 특징을 생성하는 과정을 가졌던 것과 다르게 CNN기반의 영상분류는 입력된 영상으로부터 분류에 필요한 특징을 자동으로 추출하고 분류하는 방법을 학습한다. CNN기반의 방법은 기존의 방법들에 비해 간단한 구조와 동시에 높은 분류 정확도를 보인다. CNN 기반의 방법은 크게 합성 곱 계층(Convolution layer)[26], 배치정규화(Batch Normalization)[27], 풀링(Poling), 활성화 함수(Activation Function)의 공통된 구조를 갖는다.

2.3.1 CNN 구조

Conv계층은 영상을 $n \times n$의 필터 단위로 영상을 순회하면서 각 부분에서 세밀한 특징을 찾는 방법이다. 예를 들어 그림 2.10-(a)는 고주파 필터의 한 종류이다. 그림 2.10-(b)의 3×3영상이라고 할 때, (b)의 값은 그림 2.10-(a)의 Conv을 적용한 후의 값은 주변 픽셀과 필터의 가중치를 곱해 모두 합한 값으로 바뀐다. Conv 연산은 영상 전체를 픽셀 단위로 순회하며 진행하기 때문에 영상의 크기가 변하게 된다. Conv계층 전 후로 영상의 필터의 사이즈에 따라 감소한다.

$$O = \frac{I + 2p - k}{s} + 1 \qquad (2\text{-}9)$$

필터링 후 영상의 크기는 주변에 특정한 값을 채워 영상의 크기를 증가시키는 패딩(그림 2.5)과 필터를 이동할 간격을 의미하는 스트라이드(stride)로 조절할 수 있다. 필터링 후 영상을 구하는 방법은 식 (2-9)과 같다. 식(2-9)에서 O는 필터링 후의 영상의 크기, I는 입력 영상의 크기, p는 패딩의 크기, k는 필터의 크기, s는 스트라이드의 크기를 의미한다.

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \quad \begin{bmatrix} 50 & 50 & 50 \\ 50 & 100 & 50 \\ 50 & 50 & 50 \end{bmatrix} \quad [200]$$

(a) (b) (c)

그림 2.10 합성 곱 연산의 적용. (a)고역통과 필터, (b)원본 영상, (c)Conv 적용 후의 행렬

배치정규화는 계층의 값의 정규화 연산을 의미한다. 신경망의 학습은 신경망의 예측 결과와 정답사이의 오차를 경사하강 알고리즘(Gradient Descent Algorithm)[28-30]을 이용해 줄여나가는 방식을 사용한다. 신경망의 깊이(계층의 수)가 깊어지면 계층의 최적화를 위한 손실함수의 가중치(Gradient)가 사라지는 현상이 발생할 가능성이 생긴다. 배치정규화는 경사소멸현상(Gradient Vanishing)을 방지한다. 또한 학습시마다 출력 값을 정규화하기 때문에 신경망의 초기 값에 대한 의존성을 줄일 수 있는 장점이 있다. 배치정규화는 전체의 데이터 중 미니배치의 B개의 입력을 정규화 한다. B개의 미니배치에 대한 입력 값의 평균과 분산은 식(2-10), (2-11)과 같이 정의되고 배치정규화를 적용한 값은 식(2-12, 2-13)로 정의된다. 식(2-13)에서 γ와 β는 각각 스케일과 이동에 관한 파라미터이다.

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \quad (2\text{-}10)$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x-\mu_B)^2 \quad (2\text{-}11)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 - \epsilon}} \quad (2\text{-}12)$$

$$y_i = \gamma\hat{x}_i + \beta \quad (2\text{-}13)$$

풀링은 영상의 크기를 줄여 작은 스케일의 영상을 만드는 방법이다. 영상의 스케일에 따라서 다양한 특징을 추출할 수 있다. 풀링은 최댓값 풀링과 평균값 풀링의 두 가지 방법을 사용한다. 최댓값 풀링은 $n \times n$의 필터사이즈에서 가장 큰 값을 출력하고 평균값 풀링은 $n \times n$의 필터사이즈의 픽셀들의 평균값을 출력한다. 그림 2.11은 임의의 행렬에 대한 최댓값 풀링과 평균값 풀링의 비교이다.

2	100	100	100
100	200	100	100
150	150	200	200
100	100	200	200

(a)

200	100
150	200

(b)

100.5	100
125	200

(c)

그림 2.11 최댓값 풀링과 평균값 풀링의 비교. (a)풀링 적용 이전의 행렬, (b)최댓값 풀링, (c)평균값 풀링

배치정규화에서 언급한 것과 같이 신경망의 학습은 경사하강 알고리즘을 이용해서 이루어진다. 하지만 고차원의 데이터는 선형함수로는 가중치의

학습이 불가능하다. 활성화 함수는 신경망에 결과에 비선형성을 부여해주는 역할을 한다. 활성화 함수를 사용하여 신경망에 비선형성을 입력과 출력의 관계가 복잡해지기 때문에 경사소멸 현상이 잘 발생하지 않게 되고 신경망의 학습이 잘 진행된다. 활성화 함수는 입력 값으로 들어온 뉴런이 활성화함수의 기준에 부합하면 다음 계층으로 값을 전달하고 부합하지 않으면 값을 전달하지 않는다. 대표적인 활성화함수로는 그림 2.12-(a)Sigmoid, 2.12-(b)Tanh, 2.12-(c)ReLU 등이 있다.

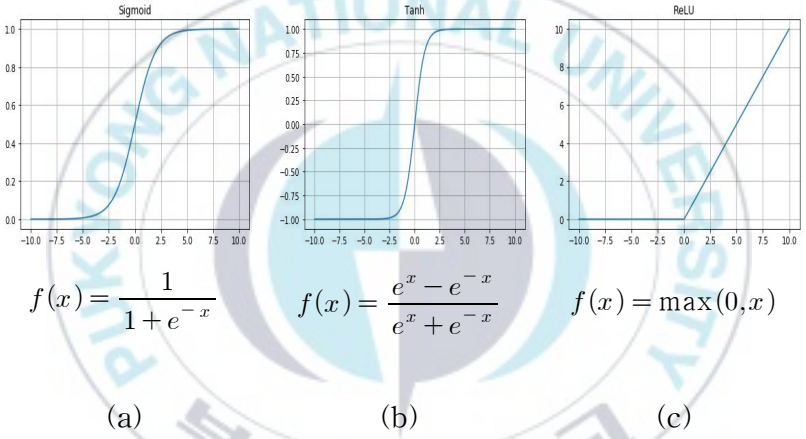


그림 2.12 대표적인 활성화 함수. (a)Sigmoid, (b)Tanh, (c)ReLU

2.3.2 CNN구조의 깊이와 성능의 상관관계

AlexNet은 2013년 1000개 이상의 카테고리에 대해 100만장의 영상을 분류하는 ImageNet 대회에서 CNN기반의 딥러닝 구조를 이용해 높은 성과를 거두었다. AlexNet 이후 많은 컴퓨터비전의 분야에서 CNN구조를 이용해 문제를 해결했다. AlexNet은 5개의 CNN계층과 3개의 완전연결계층을 이용해 네트워크를 구성했다. AlexNet은 $224\times224\times3$의 컬러영상을 입력으

로 받아 1000개의 클래스에 대해 분류한다. AlexNet의 특징은 2개의 GPU로 병렬연산을 위한 병렬적인 구조의 구성과, 구조의 과적합(over-fitting)을 막기 위해 한 장의 영상을 회전하거나 일부를 잘라내어 데이터를 다양화 하는 방법을 사용해 데이터의 수를 증가시켰다.

이후 CNN기반의 네트워크 연구는 데이터의 깊이에 따른 성능의 변화에 관한 연구가 주를 이루었다. VGG[31]는 네트워크의 깊이에 따른 성능 변화를 관찰하기 위해 커널의 사이즈를 3×3으로 고정했다. 작은 사이즈의 커널은 큰 사이즈의 커널을 분할해 실행한 것과 같다. 예를 들어 3×3의 커널은 5×5의 커널을 2번에 나누어 실행한 것과 같고 7×7의 커널을 3번에 나누어 실행한 것과 같다. 커널의 사이즈가 작을수록 Conv계층을 여러 번 실행할 수 있기 때문에 네트워크의 비선형성이 증가해 성능 향상에 유리하다.

VGG는 AlexNet과 마찬가지로 ImageNet의 데이터(224×224×3)를 이용해 총 6개의 구조를 만들어 성능을 비교했다. 각 구조에 A, A-LRN, B, C, D, E로 깊이에 따라 분류했다.

NetWork 구분					
A	A-LRN	B	C	D	E
11층	11층	13층	16층	16층	19층
Input(224×224, RGB Images)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
SOFTMAX					

표 2.3 VGG에서 사용한 네트워크의 구조들.

VGG의 구조는 표 2.3와 같다. 표 2.3에서 E를 VGG19라고 부른다. VGG19은 ImageNet에서 약 93%정도로 높은 정확도를 보였고 CNN기반의 네트워크에서 구조의 깊이가 증가할수록 성능 향상을 기대할 수 있음을 보였다. VGG는 19개의 CNN계층을 사용해 깊은 네트워크를 구성했지만 성

능의 향상을 위해 깊이가 더 깊은 네트워크를 구성하는데는 한계가 있었다. 네트워크의 깊이가 깊어질수록 기울기 소멸현상(Gradient Vanishing)이나 과적합 등의 문제가 발생할 가능성이 생긴다. VGG19에서는 파라미터의 수에 따른 학습의 어려움을 해소하기 위해 완전연결 계층의 가중치의 초기 값을 A의 11개의 계층의 학습 값으로 주어 VGG19모델을 학습했다.

ResNet[32]은 152개의 Conv계층을 이용해 네트워크를 구성했다. ResNet은 기울기 소멸현상의 해결법과 함께 깊은 네트워크에서 발생할 수 있는 Degradation 문제를 지적했다. Degradation은 네트워크가 깊어지면 성능이 얕은 네트워크보다 감소하는 현상이다. ResNet은 CNN계층과 완전연결계층을 이용해 각각 20개와 56개의 계층을 가지는 두 네트워크를 구성해 깊이에 따른 성능을 비교한 실험의 결과에서 학습데이터와 훈련 데이터 모두 계층의 수가 더 많은 네트워크가 더 낮은 성능을 보였다. ResNet은 이 결과를 바탕으로 더 깊은 네트워크를 구성하기 위한 방법으로 Residual Learning을 제안하였다. Residual Learning은 Conv계층의 출력에 입력 값을 더하는 지름길을 추가한 구조이다. 기존 Conv계층의 구조는 입력(x)를 목표 값(y)으로 변환하는 함수 $H(x)$를 얻는 과정이다. Residual Learning은 그림 2.13과 같이 $F(x)+x$를 최소화하는 변환이다. x는 입력 값이므로 변화할 수 없기 때문에 $F(x)$를 0에 가깝게 최소화한다. $F(x)$의 최소화는 $H(x)-x$를 최소화하는 것과 같고 $H(x)-x$를 잔차(Residual)이라고 한다. Resnet 구조는 기존의 Conv구조에서 간단한 지름길을 추가해 분류 성능을 높였다.

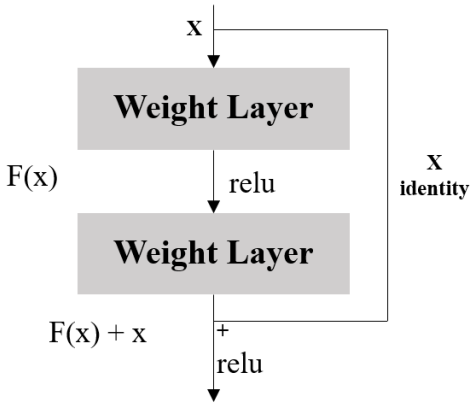
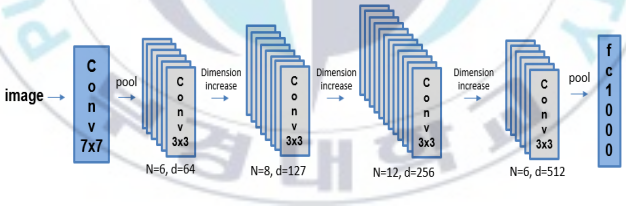
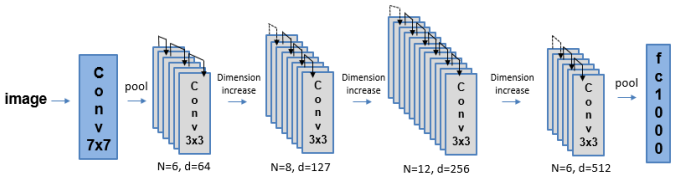


그림 2.13 Residual 구조.

ResNet은 VGG의 네트워크를 바탕으로 그림 2.14의 Plain구조와 Residual구조를 구성해 학습을 진행한 후 18개의 계층을 가진 구조의 결과와 비교했다.



(a)



(b)

그림 2.14 VGG구조 기반의 Plain구조와 Residual구조. (a)Plain 구조, (b) Residual구조

실험의 결과에서 그림 2.14-9(a)는 34개의 CNN계층을 가지지만 Degradation으로 인해 기존의 18개 계층의 계층을 가지는 구조의 성능보다 감소한 결과를 보였고 Residual은 Plain구조와 동일한 계층을 수를 가지고 있지만 성능이 증가한 결과를 보였다. ResNet의 결과는 Identify Mapping을 통해 깊은 구조에서의 Degradation 문제와 기울기 소멸현상을 해결할 수 있음을 보였다. 이후 많은 CNN기반의 구조에서 Identify Mapping을 이용해 구조를 설계한다.

[20]은 ResNet의 구조를 스테그아날리시스에 도입했다. [20]에서는 Identify Mapping의 구조로 그림 2.15-(a)의 비병목블록과 그림 2.15-(b)의 병목블록을 사용했다. 그림 2.15-(a)의 경우 ResNet에서 사용한 구조와 동일하고 그림 2.15-(b)의 경우 1x1의 Conv계층을 추가해 1개의 구조에 3개의 Conv계층을 사용한다. [20]은 그림 2.9-(b)의 한 개의 고역통과필터를 적용한 후 Residual구조를 추가해 깊이에 따른 커버영상과 스테고영상의 분류 정확도를 분석했다. 그림 2.16와 같이 50개의 Conv계층을 가지는 네트워크를 구성했을 때 WOW의 경우 95.7%, UNIWARD의 경우 93.7%의 분류 정확도를 보였다.

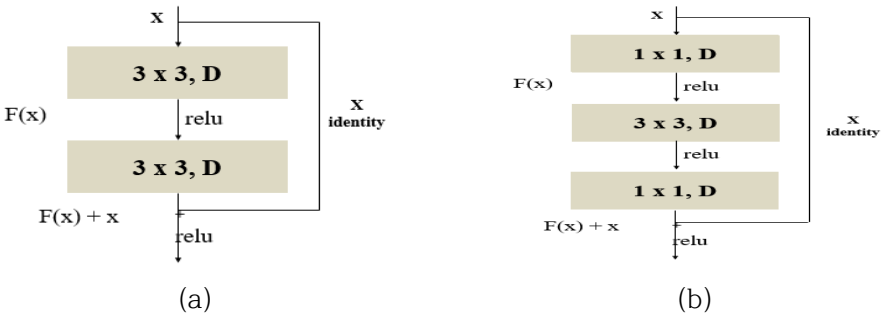


그림 2.15 [20]에서 사용한 Residual 블록. (a)비병목 Residual블록, (b)병목 Residual블록

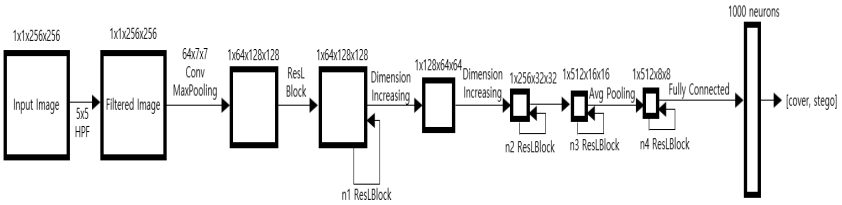


그림 2.16 [20]의 네트워크 구조.

그 외 CNN기반의 스테그아날리시스 방법들은 Xu의 구조에서 고역통과 필터의 수를 3개로 증가시킨 Multi-Channel[33], 전이학습을 이용해 서로 다른 3가지 Sub-Network를 학습시킨 Rest-Net[34], 공분산 풀링을 이용해 학습 속도와 정확성을 증가시킨 [35]와 같이 구조의 다른 부분들의 변경에 따른 분류 성능의 변화를 관찰했다.

2.3.3 ResNet 구조 기반의 방법들

ResNet의 영상 분류에 있어서 큰 성공 이후 많은 ResNet 기반의 방법들이 연구되었다. Preact-ResNet[36]은 Resnet 블록을 구성을 변경해 가며 성능을 분석했다. 기존의 Resnet 블록은 그림 2.17과 같이 Conv, BN, ReLU, Conv, BN 후 출력 값을 입력에 ReLU를 적용해 Identifying Mapping을 하는 구조를 가졌다. [36]에서는 기존의 Residual구조를 변형해 성능의 변화를 분석했다. 그림 2.17-(b) ~ (g)는 그림 2.17-(a)를 기반으로 변형된 구조들이다.

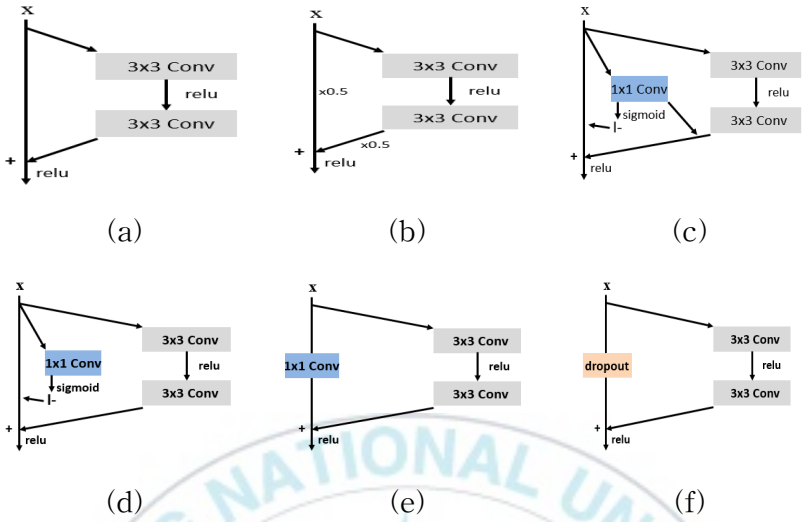
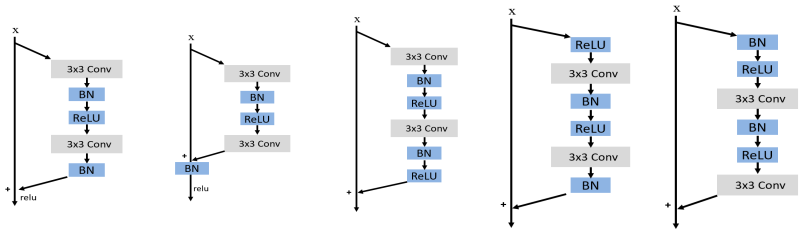


그림 2.17 변형 Residual 구조. (a)original, (b)Constant Scaling, (c)Exclusive gating, (d)shorcut-only gating, (e)conv shortcut, (f)dropout shortcut

[36]에서는 그림 2.17과 같이 Resnet블록을 변경하면서 성능의 변화를 관찰하였으나 기존의 구조인 그림 2.17-(a)에서 가장 좋은 결과를 보였다. [Pre]에서는 이러한 결과의 원인으로 Resnet 블록의 변화는 구조의 표현력을 증가시킬 수 있으나 이에 따른 학습 난이도의 증가로 인한 성능의 하락으로 추정했다.

[36]에서는 Resnet 블록의 변형과 더불어 Resnet 블록에서 활성화 함수의 위치에 따른 성능을 분석했다. 그림 2-18은 각 Residual 구조와 구조에 따른 학습 오차이다.

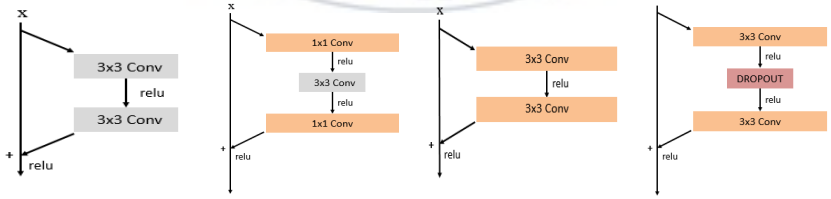


(a) Error : 6.61%　(b) Error : 8.17%　(c) Error : 7.84%　(d) Error : 6.71%　(e) Error : 6.37%

그림 2.18 [36]의 변형 Residual 구조. (a)original, (b)BN after addition, (c)ReLU before addition, (d)ReLU-only pre-activation, (e)full pre-activation

그림 2.18의 결과에서 기존의 그림 2.18-(a)구조에 비해 배치 정규화와 활성화 함수를 Conv 계층보다 먼저 적용한 그림 2.18-(e)의 full pre-activation에서 가장 좋은 성능을 보였다. 그림 2.18-(e)의 결과는 정규화된 입력이 Conv 계층에 전달되기 때문에 분류 성능이 더 상승한다.

Wide-ResNet[37]은 ResNet의 성능을 높이기 위해 구조의 깊이뿐만 아니라 구조의 넓이(특징 맵의 수)에 따른 성능의 변화를 관찰했다.

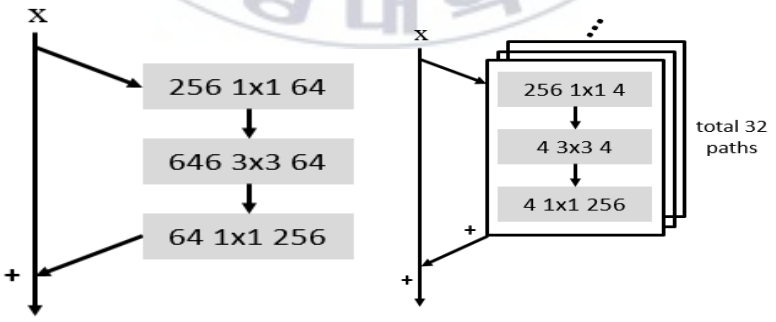


(a)　(b)　(c)　(d)

그림 2.19 [37]의 Residual 구조. (a)Original, (b)BottleNeck, (c)Original-Wide, (d)wide-dropout

[37]은 그림 2.19-(a)와 같은 기존의 Reisdual 구조에서 그림 2.19-(b) ~ (d)와 같이 Residual 구조의 넓이(필터의 수)를 증가시켜 가며 성능을 분석했다. [37]은 기존의 ResNet의 결과와 검증 오차는 비슷한 수준을 보이나 학습에 소요되는 시간을 대폭 감소시켜 더 효율적인 학습을 가능하게 했다. [37]의 결과를 바탕으로 CNN 네트워크의 성능 향상을 위해서는 구조의 깊이 뿐 아니라 구조의 넓이도 고려할 필요가 있음을 알 수 있다.

이전까지의 ResNet기반의 방법들은 CNN 네트워크의 깊이와 넓이의 변화에 따른 성능의 변화를 주로 연구했다. ResNext[38]에서는 ResNet의 구조를 기반으로 Residual 블록에 Cardinality라는 새로운 차원의 개념을 도입하였다. Cardinality는 기존의 Residual 블록의 특징 맵을 작은 단위로 나누어 병렬적으로 처리하는 방식을 의미한다. 그림 2.20-(a)는 기존의 Residual 블록이다. [38]에서는 (b)와 같이 256개의 특징 맵을 4개씩 32개의 그룹으로 나누어 처리한 후 마지막에 합해주는 방법을 사용했다. Cardinality의 개념을 사용한 ResNext의 구조는 기존의 ResNet-50과 거의 비슷한 연산량을 가지면서 성능은 더 우수한 결과를 얻었다.



(a) (b)

그림 2.20 [38]의 Residual 구조. (a)original, (b)ResNext cardinality = 32

ResNet은 Residual 블록에서는 입력 특징 맵과 출력 특징 맵의 수를 동일하게 유지하고 영상의 크기를 줄이는 계층에서 특징 맵의 수를 두 배로 증가시키는 방법을 사용한다. PyramidNet[39]은 Residual구조의 성능이 다운샘플링과 특징 맵의 수의 증가 과정에 큰 영향을 받는다는 점에 착안해 특징 맵 수의 증가를 CNN 네트워크 전체 계층에 적용하는 방법을 도입했다. ResNet이 Residual 블록에서는 특징 맵의 수가 일정하게 유지된 반면, [39]는 Residual 블록에서도 특징 맵의 수가 증가하는 구조를 도입했다. Residual 블록에서 특징 맵의 수가 증가할 경우 Identify Mapping을 하는 과정에 입력 특징 맵의 수와 출력 특징 맵의 수에 불일치가 발생한다. [39]는 zero-padding을 이용해 입력 특징 맵의 수와 출력 특징 맵의 수를 일치시켜 Identify Mapping을 적용했다. [39]는 특징 맵의 수를 증가시키는 방법으로 선형적(linear) 증가와 기하급수적(exponential) 증가의 두 가지 구조를 통해 비교했다. [39]의 결과는 선형적으로 특징 맵의 수를 증가시키는 것이 더 좋은 결과를 보였고 기존의 Resnet 구조보다 더 높은 성능을 얻었다.

2.3.4 DenseNet 구조

DenseNet[40]은 입력의 특징 맵을 출력의 특징 맵에 더해준 ResNet기반의 방법들과 달리 이전 출력 계층의 특징 맵을 다음 계층의 입력에 연결해주는 방식을 사용했다. DenseNet의 직접 연결 구조는 앞 계층의 연산 결과를 직접 뒷 계층으로 전달한다. 이전 계층의 특징 맵을 재사용해 학습에 필요한 파라미터의 수를 줄일 수 있다(그림 2.21).

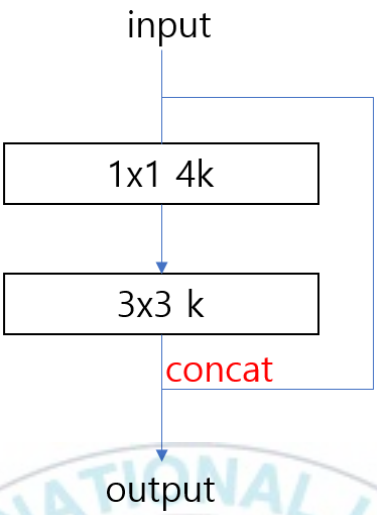


그림 2.21 Dense Block의 구조.

DenseNet 구조는 이전 계층의 특징 맵을 다음 계층으로 연결해주기 때문에 각 계층의 특징 맵의 수가 클 경우 전체 구조의 특징 맵의 수가 지나치게 커질 수가 있다. 따라서 DenseNet 구조에서는 각 계층의 특징 맵의 수를 작은 값을 사용한다. DenseNet에서 계층의 특징 맵의 수는 growth rate(k)에 따라 결정된다. DenseNet는 병목구조를 사용하고 k를 4배씩 증가시켰을 때 가장 좋은 결과를 보였다. 그림 2.22는 DenseNet 기반의 네트워크 구조이다. DenseNet은 ResNet보다 상대적으로 가벼운 네트워크를 구성했으나 네트워크 앞단의 병목구조에서 연산량이 크게 증가와 growth rate를 4배로 설정에 대한 문제가 있었다.

2.3.5 PeleeNet 구조

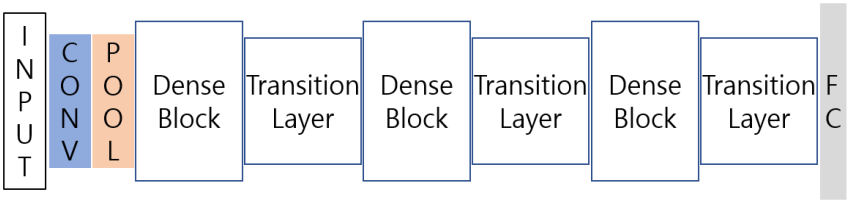


그림 2.22 DenseNet 기반의 네트워크.

PeleeNet[41] 구조는 DenseNet 구조를 기반으로 DenseNet구조보다 더 빠르고 높은 성능을 보이는 구조를 제안했다. PeleeNet은 기존의 DenseNet 구조의 초기 병목구조에서 입력 특징 맵의 수가 병목구조의 특징 맵의 수가 줄어들어 구조에서 전체 네트워크의 연산량이 증가하는 문제를 개선하기 위해 DenseNet의 Dense블록을 2개의 방향을 가지는 구조로 변경하고, DenseNet의 transition 계층을 Stem블록으로 변경했다.

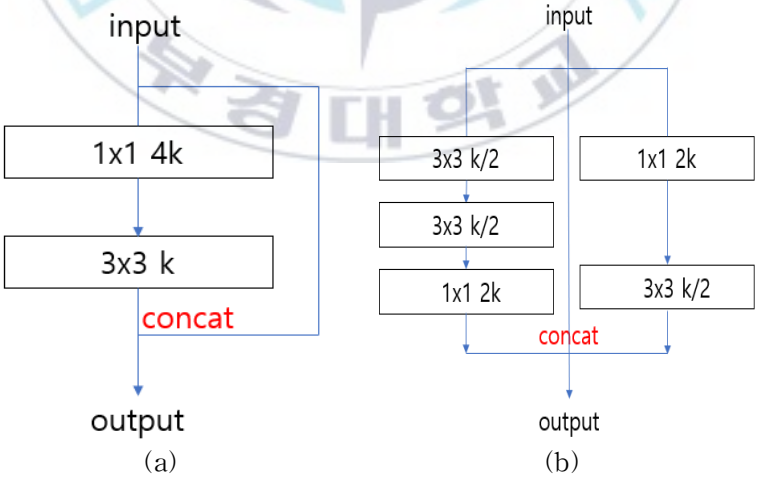


그림 2.23 DenseNet과 PeleeNet의 Dense블록. (a)기존 Dense Block, (b)[41]의 Dense Block

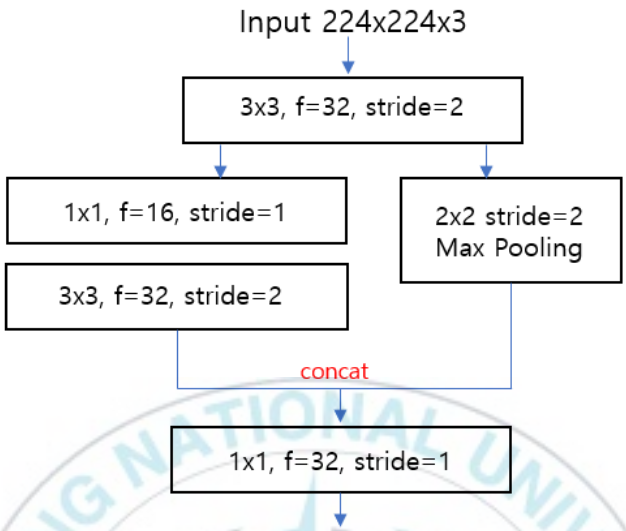


그림 2.24 [41]의 Stem Blcok.

그림 2.23는 각각 기존의 Dense블록과 PeleeNet에서 사용한 Dense블록이다. PeleeNet의 블록은 기존의 구조에서 1개의 방향을 추가해 DenseNet의 급격한 연산량의 증가 문제를 해결했다. 그림 2.24는 DenseNet의 transition계층을 변형한 Stem블록 구조이다. 기존의 transition계층은 특징맵의 수를 두 배로 증가시키고 영상의 크기를 1/4로 줄이는 연산을 했다. Stem 블록은 입력으로 224x224의 컬러 영상이 들어왔을 때 출력은 가로 세로를 4배 줄인 32채널의 영상이 된다. Stem블록은 조금의 연산량을 추가해 모델의 표현 성능을 높이는 역할을 한다. 그림 2.25는 PeleeNet 기반의 네트워크이다. PeleeNet은 변형된 Dense블록과 Stem블록을 통해서 기존의 DenseNet 구조의 문제점을 개선하고 큰 성능 향상을 얻었다.

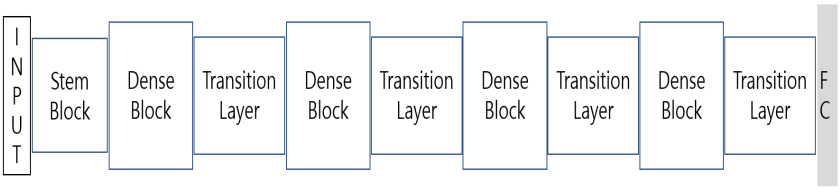


그림 2.25 [41]기반의 네크워크.

Xu, VGG와 같은 초기 CNN 기반의 네트워크들은 구조의 깊이에 따른 성능을 분석해, 각 문제에 적합한 CNN구조의 최적화에 초점을 두었다. 그 이후 Resnet, DenseNet과 같은 방법들은 입출력 전달 방식의 변화를 통한 성능을 향상시켰고 또한 네트워크의 최적화를 통해 초기의 구조들보다 훨씬 더 깊은 구조를 사용해 구조의 성능을 비약적으로 향상시킬 수 있었다. PeleeNet은 DensetNet과 같은 입출력 전달 방식을 사용했지만 네트워크의 크기를 감소시키면서 성능은 더 좋은 결과를 보였다.

2.4 복수의 스테가노그래피 알고리즘 분류

기존의 스테그아날리시스는 데이터가 삽입되지 않은 커버영상과 데이터가 삽입된 스테고영상을 분류하는 문제를 다뤘다. 스테그아날리시스 문제에 CNN구조 기반의 방법이 도입된 이후, 커버영상과 스테고영상을 분류하는 스테그아날리시스의 분류 정확도는 상당히 높은 수준을 보인다. 하지만 기존의 스테그아날리시스는 데이터의 삽입 여부를 검출할 뿐, 스테고영상에 삽입된 데이터를 검출하는 일은 요원하다. 본 논문에서는 검출된 스테고영상에서 삽입된 데이터를 검출하기 위한 첫 번째 단계로 복수의 스테가노그래피 알고리즘들을 분류한다. 기존의 스테그아날리시스 문제와 달리 본 논문에서의 스테그아날리시스 문제는 스테고영상을 하나의 클래스가 아

닌 각 스테가노그래피 방법에 대한 각각의 클래스로 분류한다. 따라서 분류의 결과는 커버영상과 스테고영상이 아닌 커버영상과 각 클래스의 분류를 수행한다.

이전의 [21]에서는 COVER, WOW, UNIWARD의 3진 분류를 통해 복수의 스테가노그래피를 분류를 위한 네트워크의 구성과 CNN기반의 스테그아날리시스 구조에서 복수의 스테가노그래피를 분류정확도에 영향을 주는 다양한 요인들에 대해서 분석하였다. WOW와 UNIWARD와 같이 정교한 스테가노그래피 알고리즘은 커버영상과 통계적으로 유사한 분포를 보인다. 데이터의 삽입으로 인한 영상의 변화 정도를 왜곡함수로 정의하고 왜곡도가 낮은 픽셀에만 데이터를 삽입하는 유사한 방법을 사용하기 때문에 스테고영상의 검출정확도가 상대적으로 낮다.

학습 데이터	검증 데이터	정확도(%)			
		COVER	UNIWARD	WOW	AVERAGE
UNIWARD	UNIWARD	84.02	73.56	-	78.80
WOW	WOW	77.63	-	78.25	77.94
UNIWARD	WOW	84.02	-	60.48	72.25
WOW	UNIWARD	77.64	67.13	-	72.38

표 2.4 WOW와 UNIWARD 교차 검증

표 2.4에서 학습 데이터는 CNN 구조를 학습시킨 데이터, 검증 데이터는 학습된 구조를 검증한 데이터를 의미한다. 학습 데이터와 검증 데이터가 동일할 경우, UNIWARD는 평균 78.8%, WOW는 평균 77.94%의 정확도를 보였다. WOW와 UNIWARD의 데이터를 교차해 검증했을 때 UNIWARD로 학습한 구조는 WOW의 데이터를 60.48%, WOW로 학습한 구조는 UNIWARD의 데이터를 67.13%를 각각의 스테고영상으로 인식했다. 스테가노그래피 알고리즘들의 유사성은 복수의 알고리즘 분류를 할 때 알고리

즘 간의 오 분류가 발생할 가능성이 있다.

[21]에서는 대부분의 스테그아날리시스 방법들에서 사용되는 고역통과필터를 이용한 영상의 전처리 과정을 분석했다. SRM은 30개의 선형, 비선형의 방향성 고역통과필터를 이용해 영상을 전처리하는 과정을 사용한다. [21]의 COVER, WOW, UNIWARD의 3진 분류의 실험 결과에서, 30개의 SRM필터를 모두 사용한 결과보다 SRM필터 중 효율적인 10개의 필터를 사용했을 때 평균 분류 정확도가 크게 증가했다. 또한 CNN구조에 따라서 평균 분류 정확도가 크게 변화하는 것을 확인할 수 있었다.

본 논문에서는 CNN구조를 통한 기존의 2진 분류와 [21]의 3진 분류 결과를 참고하여 COVER, LSB, PVD, WOW, UNIWARD, MIPOD의 6개의 클래스를 분류한다. 복수의 스테가노그래피 알고리즘의 분류는 기존의 스테그아날리시스는 스테가노그래피 알고리즘에 관계없이 커버, 스테고로 2진 분류한다. 커버, 스테고 2진분류의 경우 영상에 적용된 스테가노그래피 알고리즘에 따라 스테고 영상을 커버영상으로 분류할 가능성이 있다. 따라서 이러한 오분류를 방지하기 위해서는 스테가노그래피 알고리즘들을 분류할 필요가 있다.

CNN구조 기반의 스테그아날리시스에서 분류하려는 클래스의 수가 증가할수록 평균적인 정확도는 감소하는 경향을 보인다. 이는 스테가노그래피의 적용으로 인한 영상의 변화가 매우 작고, 정교한 스테가노그래피 알고리즘일수록 스테고영상의 커버영상과 유사한 분포를 가지기 때문이다. 따라서 복수의 스테가노그래피 분류 정확도 개선을 위해서는 스테가노그래피의 알고리즘들의 특성을 이용해 작은 차이를 분별할 수 있어야 한다. 제안 방법에서는 적용된 스테가노래피 방법에 따른 스테고 영상들 간의 계층적 특징을 이용해 복수의 스테가노그래피를 분류하는 계층적 CNN 네트워크를 사용한다.

Ⅲ.계층적 CNN 구조의 스테그아날리시스

스테가노그래피는 데이터의 삽입으로 인한 영상의 왜곡 정도를 최소화하는 것을 목표로 한다. 스테가노그래피의 적용으로 인한 영상의 왜곡은 매우 작고, 스테가노그래피 알고리즘 간의 차이도 미소하다. 복수의 스테가노그래피 알고리즘을 분류하기 위해서는 커버영상과 스테고영상 뿐 아니라 스테가노그래피 알고리즘 간의 작은 차이를 분류할 수 있어야 한다.

복수의 스테가노그래피 알고리즘을 분류하는 방법은 기존의 다른 분류 문제들과 마찬가지로 CNN 구조의 출력 클래스의 수를 증가시키는 방법이 있다. 하지만 스테그아날리시스 문제는 스테가노그래피 클래스 사이의 작은 차이를 분류하기 때문에 출력 클래스의 증가는 네트워크의 분류 기준이 복잡해져 전체적인 분류 성능의 감소를 야기할 가능성이 있다. 또한 단일 네트워크를 통한 분류는 출력의 수가 증가함에 따라 구조를 학습하기 위해 각 클래스의 데이터를 동시에 포함하기 때문에 학습에 필요한 하드웨어 요구와 학습시간이 증가한다.

본 논문에서는 복수의 스테가노그래피 알고리즘을 효과적으로 분류하기 위한 계층적 CNN 구조 기반의 스테그아날리시스 방법을 제안한다. 제안 방법에서는 기존의 단일 네트워크 기반의 복수의 스테가노그래피 알고리즘 분류의 한계인 높은 하드웨어 성능의 필요와 학습시간의 증가 문제를 해결하기 위해 복수의 네트워크를 이용해 클래스를 계층적으로 분할, 분류하는 계층적 CNN 구조를 사용한다. 계층적 CNN구조는 스테가노그래피 알고리즘들의 계층적인 특성을 이용한다. 예를 들어 PVD로 충분히 학습된 구조는 LSB, WOW, UNIWARD, MIPOD과 같은 다른 스테가노그래피 데이터를 스테고영상으로 분류하지 못하는데, 이러한 계층적 특징을 이용해

계층적으로 스테가노그래피 알고리즘을 분류한다.

계층적 CNN구조는 복수의 알고리즘 분류를 나누어 함으로써 학습에 필요한 자원을 줄일 수 있고, 각각의 층에서 다른 구조와 분류 방법을 사용해 최적화된 구조를 설계할 수 있는 장점이 있다. 본 절에서는 CNN 기반의 스테그아날리시스 네트워크의 구성과 네트워크의 학습을 위해 사용한 하이퍼 파리미터에 대해 설명한다.

3.1 하이퍼 파라미터

CNN 기반의 네트워크의 성능은 네트워크의 구조, 사용한 데이터의 품질, 학습 횟수, 배치의 크기 등 다양한 요소에 영향을 받는다. 제안 방법에서 사용하는 구조들의 정량적인 분석을 위해 각 구조의 특징을 제외한 옵티마이저, 학습 횟수, 학습률, 배치사이즈 등의 하이퍼 파라미터는 동일한 값을 사용한다. 구조에서 사용한 하이퍼 파라미터의 값은 아래 표3.1과 같다.

파라미터	값	비고
epoch	100	
batch size	10	
learning rate	0.001	매 10 epoch마다 90%로 감소
Momentum	0.9	가중치 갱신 방법

표 3.1 제안 방법에서 사용한 하이퍼파리미터.

표 3.1에서 epoch은 모든 구조가 모든 데이터를 한번 씩 학습하는 것을 말한다. 각 구조는 Cover/Stego 데이터를 전체를 총 100번 학습하는 과정

을 가진다. 배치사이즈는 학습 단계에서 가중치를 갱신하기 위해 한 번에 참조하는 데이터의 수를 의미한다. 총 30,000개의 데이터가 1 epoch일 때 30,000의 데이터를 10개씩 무작위로 가져와 구조의 가중치를 갱신한다. 배치사이즈가 증가할수록 학습에 걸리는 시간이 감소하고 가중치를 갱신할 때 참조되는 데이터가 많아지기 때문에 지기 때문에 분류성능이 증가하는 경향이 있다고 한다. 따라서 일반적으로 배치사이즈는 가용한 큰 값을 사용하지만, 본 논문에서 사용한 구조 중 ResNet 구조는 네트워크의 구성에 필요한 메모리가 커 사용가능한 최대 배치사이즈가 10이였다. 또한 구조에 따라서 작은 배치사이즈(10)를 사용했을 때 더 좋은 성능을 보이는 경우도 있었다. 각 구조의 실험결과의 동일한 비교를 위해 모든 구조에서 10의 배치사이즈를 사용했다.

학습률은 가중치가 갱신되는 정도를 의미한다. 학습률이 커질수록 가중치가 변하는 정도가 커지기 때문에 학습의 속도가 증가하지만 동시에 수렴이 어려워진다. 일반적으로 학습률은 특정한 값에서 시작해서 학습이 진행됨에 따라 학습률을 줄여 결과의 수렴이 잘 되도록 하는 방법을 사용한다. 학습률은 단계적으로, 선형, 비선형 등 감소하는 비율에 다양한 방법이 있다. 본 논문에서 사용한 학습률 감쇄의 방법은 Cosine Deacy를 사용한다. Cosine Decay는 학습률을 코사인 함수의 분포에 따라 감소시키는 방법으로, 학습 초기에는 학습률이 빠르게 감소하고 후에 갈수록 감쇄 속도가 줄어드는 특징이 있다. 제안 방법의 학습률은 처음에 0.001에서 시작해 매 10번의 epoch마다 90%로 감소한다.

3.2 계층적 CNN 네트워크 구성

CNN 기반의 분류문제에서 네트워크의 구조는 성능에 가장 큰 영향을 미친다. AlexNet과 같은 초기의 CNN 구조 이후, 네트워크의 구성과 특징 전달 방법에 대해 다양한 연구가 진행되었다. 일반적으로 CNN 기반의 구조는 Conv계층과 배치 정규화, 활성화 함수, 풀링으로 구성된다. 입력으로 들어오는 한 장의 영상에 대해 Conv 계층과 배치 정규화, 활성화 함수를 통해 특징 맵(채널)의 가중치의 수를 증가, 변형하는 과정을 가진다. 그 이후 풀링을 통해 영상의 크기를 줄이고 최종적으로 N개의 특징 맵을 가진 1x1의 영상을 생성한다. 특징 맵과 완전연결 계층의 선형 연산을 통해 n개의 클래스에 대한 각각의 확률 값을 도출해 분류 구조를 학습한다. 활성화 함수와 풀링 계층의 경우 다양한 방법들이 제안되었으나 주로 ReLU와 평균 값 풀링을 사용한다. 추출된 특징을 전달하는 방법의 경우 앞 절에서 언급한 것과 같이 ResNet과 같이 입력 값을 출력 값에 더해주는 방법, DenseNet, PeleeNet과 같이 입력 특징 맵을 출력 특징 맵에 직접 연결해 주는 방법 등이 있다.

제안 방법에서는 Xu기반의 Plain구조(PP), Plain구조에 ResNet구조의 개념을 도입한 DRL구조(DRL), ResNet구조, DensNet구조, PeleeNet구조를 통해 구조에 따른 성능을 비교한다. Xu기반의 구조에 CNN계층의 수를 1개더 증가시킨 것을 제외하면 2.3.3 ~ 2.3.5절에서 언급한 구조들에서 고역통과필터의 적용을 제외하고 동일한 구조를 사용한다.

본 논문에서는 복수의 스테가노그래피 알고리즘의 분류를 위해 동시 분류와 계층 분류의 방법을 비교한다. 동시 분류는 6개의 클래스를 동시에 분류하는 방법이고 계층 분류는 2.1.6에서 언급한 스테가노그래피 알고리즘의 특성을 이용해 각 클래스를 순차적으로 분류하는 방법이다. 계층 분류의

경우 PVD, LSB, UNIWARD의 순서로 데이터를 검출하고 최하위 계층에서 COVER, WOW, MIPOD을 분류한다. 계층적 구조와 단일 구조의 비교는 아래 그림 3.1과 같다.

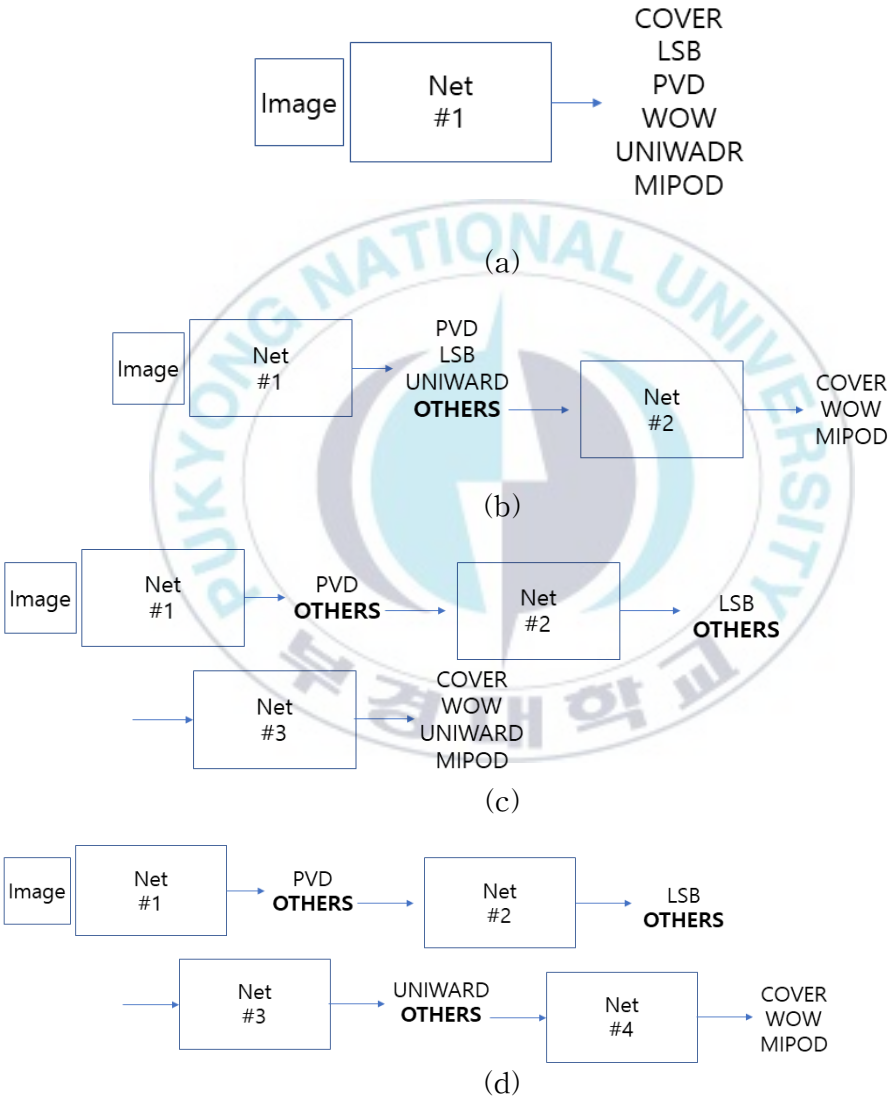


그림 3.1 단일 네트워크와 제안방법의 계층적 네트워크 구조. (a)단일 네트워크, (b)2계층 네트워크, (c)3계층 네트워크, (d)4계층 네트워크

그림 3.1(a)는 입력 영상에 대해서 각 클래스에 해당하는 확률 값을 도출한다. 계층 네트워크는 복수의 네트워크를 사용해 각 층에 해당하는 특정 스테가노그래피 알고리즘의 확률과 나머지 클래스의 확률을 도출한다. 3.1-(b)는 2계층 네트워크로 상위 계층에서 PVD, LSB, UNIWARD를 검출하고 하위 계층에서 COVER, WOW, MIPOD을 분류한다. 3-1(c)는 3계층 네트워크이다. PVD, LSB의 순서로 알고리즘을 검출하고 하위 계층에서 COVER, UNIWARD, WOW, MIPOD을 분류한다. 3-1(d)는 4계층 네트워크이다. PVD, LSB, UNIWARD 순서로 데이터를 검출하고 하위 계층에서 COVER, WOW, MIPOD을 분류한다. 계층적 구조에서 2분류의 경우 동시 분류 방법보다 비슷하거나 낮은 성능을 보여 결과에서 제외했다.

PUKYONG NATIONAL UNIVERSITY
부경대학교

Ⅳ.실험 방법 및 결과

본 논문에서는 COVER, LSB, PVD, WOW, UNIWARD, MIPOD의 6개 클래스에 대해 계층적 CNN구조를 사용해 분류한다. 제안 방법의 성능 평가를 위해 기존의 단일 CNN구조 기반 6진 분류와 결과를 비교한다.

복수의 알고리즘 분류에 가장 좋은 구조를 결정하기 위해 제안 방법에서는 Xu기반 Plain구조, ResNet 기법을 사용해 Plain 구조를 변형한 구조, ResNet, DenseNet, PeleeNet을 사용했다. 각각의 구조별로 필요 메모리의 크기가 다르기 때문에, 동등한 비교를 위해 동일한 배치사이즈를 사용했다. 본 논문에서는 GeForce 1080 Ti의 11GB 그래픽 메모리를 사용했는데, ResNet 구조에서 최대 가능한 배치사이즈가 10임을 고려하여 모든 구조에서 배치사이즈 10과 100의 epoch으로 설정했다. 학습은 파이토치[43] 라이브러리를 사용해 진행했다.

4.1 동시 분류

동시 분류는 단일 CNN구조를 통해 이루어진다. 입력으로 들어온 영상은 CNN구조를 거쳐 각각의 클래스에 해당하는 확률 값을 출력한다. 표 5.1은 5개의 클래스에 대한 동시 분류의 결과이다. 각 클래스마다 30,000장의 데이터를 사용하는데, 본 논문에서 사용한 32GB의 RAM에서 최대 150,000의 데이터를 동시에 사용할 수 있었다. 따라서 MIPOD을 제외한 5개의 클래스에 대한 분류결과를 바탕으로 복수의 스테가노그래피 알고리즘 분류를

위한 최적의 구조를 결정한 후 각 클래스에서 5000장의 학습 데이터를 감소시켜 6개의 클래스에 대해 동시분류 방법을 적용했다. 표4.1의 결과에서 PVD의 경우 모든 구조에서 99% 이상의 정확도를 보였다. 1LSB는 Dense 구조를 제외하면 모두 높은 분류 정확도를 보였으나, 상대적으로 복잡한 구조를 가지는 ResNet, Pelee 구조가 Plain구조의 PP보다 낮은 분류 정확도를 보였다. WOW와 UNIWARD 알고리즘의 경우 각각 약 77%, 88%로 Pelee 구조에서 가장 높은 분류 정확도를 보이고, 평균 정확도 또한 Pelee 구조에서 가장 높은 결과를 보인다. Dense 구조의 결과에서 Cover 분류 정확도는 약 97%로 상당히 높은 결과를 보이는데, 이는 Dense구조가 다른 구조에 비해 입력 영상을 Cover영상으로 잘못 분류하는 경우가 많은데서 기인한다. 5진 분류의 결과에서 Pelee이 복수의 스테가노그래피 알고리즘을 분류하기에 최적화된 구조임을 알 수 있다. 5진 분류의 결과를 바탕으로 MIPOD을 추가한 6개의 클래스를 분류했을 때 약 78%정도의 평균 정확도를 얻었다.

	COVER	1LSB	PVD	WOW	UNIWARD	AVG
PP	0.5528	0.9738	0.9993	0.7568	0.6182	0.7802
DRL	0.6856	0.9710	0.9994	0.7631	0.5874	0.8014
Res	0.7641	0.9164	0.9976	0.7213	0.7835	0.8366
Dense	0.9745	0.8215	0.9963	0.4855	0.3655	0.7287
Pelee	0.7608	0.9278	0.9989	0.7708	0.8881	0.8693

	COVER	1LSB	PVD	WOW	UNIWARD	MIPOD	AVG
Pelee	0.6810	0.8969	0.9983	0.7105	0.8653	0.5629	0.7858

표 4.1 단일 네트워크 기반 5진 분류 및 6진 분류

동시 분류의 결과는 각 스테가노그래피 알고리즘의 커버, 스테고 2진 분류의 결과에 비해 상당히 감소된 정확도를 보이고 분류하려는 클래스의 수

가 증가할수록 평균 분류 정확도는 급격히 감소한다. 이는 스테가노그래피 알고리즘들을 분류하기 위한 네트워크의 분류기준이 너무 복잡해져 전체적인 분류 성능의 감소를 야기한다.

4.2 계층 분류

계층적 구조는 스테가노그래피 알고리즘을 계층적 특징을 이용해 알고리즘을 차례대로 검출하기 때문에, 상위 계층에서의 분류 정확도가 높다면 특정 스테가노그래피 알고리즘을 제외한 대부분의 데이터가 하위 계층으로 전달된다. 따라서 동시 분류 방법에서 발생하는 클래스 간 잘못된 분류의 가능성이 줄어들게 된다. 또한 계층 구조는 단일 구조를 사용하는 동시 분류 방법과 달리 각 계층에 다른 CNN구조를 사용해 분류하려는 대상에 최적화된 구조를 사용할 수 있는 장점이 있다.

계층의 구성은 2.1.6에서 언급한 스테가노그래피의 계층적 특징을 따라서 2계층구조, 3계층구조, 4계층구조를 구성한다.

계층구조는 각 계층의 분류 과정에서 데이터의 오분류가 누적될 가능성이 있다. 따라서 각 계층에서 특정 알고리즘의 분류 정확도는 높이고 목표가 아닌 알고리즘들은 모두 커버영상으로 분류되어야 한다. 각 계층에서 발생하는 손실을 Diff로 정의한다. Diff는 해당 계층에서 목표 알고리즘과 다른 알고리즘들의 데이터 교차 검증을 통해, Cover의 분류정확도와 목표 알고리즘의 분류정확도를 더한 값에서 나머지 알고리즘들의 정확도를 뺀 값이다. Diff의 값이 클수록 해당 계층에서 발생하는 상대적인 손실이 작다.

4.2.1 4계층 분류

4계층 분류는 4개의 Network를 사용해 데이터를 분류한다. 첫 번째 계층에서 PVD를 검출하고 두 번째와 세 번째 계층에서 각각 1LSB, UNIWARD를 검출한다. 마지막 계층에서는 COVER와 WOW, MIPOD을 분류한다. 4계층 분류에서 각 계층의 성능은 아래의 표 4.2~4.5와 같다. 표 4.2에서 PVD의 경우, Pelee 구조에서 COVER와 PVD 모두 약 99% 이상의 정확도를 보이고 다른 알고리즘들에서 발생하는 손실도 약 0.3% 이하의 작은 값을 보인다.

Type	COVER	PVD	1LSB	WOW	UNIWARD	MIPOD	Diff
Pelee	0.9975	0.9995	0.0033	0.0026	0.0026	0.0026	1.9859

표 4.2 4계층 PVD 검출.

LSB 알고리즘의 경우 COVER와 LSB의 분류는 상당히 높은 정확도를 보인다. 하지만 학습에 사용된 CNN 구조에 따라서 WOW와 UNIWARD를 LSB로 분류하는 정도가 달라진다. 표 4.3은 각각의 CNN구조에 따른 LSB 검출 결과와 WOW, UNIWARD 오분류 결과이다. LSB 분류의 경우 ResNet을 사용한 CNN구조에서 WOW와 UNIWARD의 손실률이 약 2%정도로 가장 좋은 결과를 보였다.

Type	COVER	1LSB	WOW	UNIWARD	MIPOD	Diff
Pelee	0.9826	0.9486	0.0361	0.0343	0.026	1.8348
Pelee preact	0.9825	0.9550	0.0591	0.0407	0.041	1.7967
DRL	0.9825	0.9716	0.0644	0.1225	0.033	1.7342
PP	0.9778	0.9691	0.0631	0.0424	0.043	1.7984
Resnet	0.9836	0.9513	0.0249	0.0276	0.020	1.8624
Resnet preact	0.9874	0.9333	0.0227	0.0201	0.017	1.8609

표 4.3 4계층 LSB 검출.

UNIWARD의 검출결과는 Pre-Activation을 적용한 Pelee구조를 사용했을 때 가장 좋은 성능을 보인다. Pre-Activation은 데이터를 Conv계층 적용 이전에 배치정규화를 선행하는 방법이다. Cover를 약 94%, UNIWARD를 약 92%의 정확도로 분류하고 WOW 데이터의 손실은 약 5.5%정도가 발생한다.

Type	COVER	UNIWARD	WOW	MIPOD	Diff
Pelee	0.9351	0.9211	0.0618	0.06790	1.7265
Pelee preact	0.9440	0.9238	0.0552	0.0572	1.7554
Resnet	0.9350	0.9159	0.0671	0.4883	1.2955
Resnet preact	0.9302	0.9090	0.0728	0.0723	1.6941

표 4.4 4계층 UNIWARD 검출

4계층 분류의 최하위 계층에서는 최종적으로 COVER와 WOW, MIPOD을 분류한다. 표 5.5에서 Pelee구조를 기반으로 전처리 필터수를 조절했을 때 10개의 필터와 Pre-Activation을 적용한 CNN구조에서 가장 좋은 성능을 보인다. 표 4.4~4.5의 결과에서 UNIWARD, WOW, MIPOD과 같이 COVER와 상대적으로 차이가 적은 알고리즘의 경우 다른 알고리즘들과

달리 Pre-Activation을 적용했을 때 분류 정확도가 상승되는 것을 볼 수 있다.

Type	COVER	WOW	MIPOD	AVG
Pelee preAct	0.7185	0.7059	0.6615	0.6953
ResNet preAct	0.7347	0.6837	0.5816	0.6667

표 4.5 4계층 COVER, WOW, MIPOD 분류

4.2.2 3계층 분류

3계층 분류는 PVD와 LSB를 먼저 계층분류를 진행한 다음, COVER, WOW, UNIWARD를 동시에 분류하는 구조이다. COVER, WOW, UNIWARD의 경우 PVD와 LSB에 비해 상대적으로 유사한 분포를 가진다. 따라서 분류 정확도가 상대적으로 낮고 동시에 분류할 경우 계층의 수를 줄여 각 계층에서 발생하는 오분류를 감소시킬 수 있다. 3계층 분류에서 PVD와 LSB는 2계층 분류의 결과와 동일하다.

3계층 분류에서 COVER, WOW, UNIWARD, MIPOD의 분류 결과는 표 4.6과 같다. 4계층 분류의 COVER, WOW 분류와 마찬가지로 Resnet구조 기반의 Pre-Activation구조를 적용했을 때 COVER 66%, WOW 76%, UNIWARD 89%, MIPOD 66%의 가장 좋은 결과를 보인다.

Net	COVER	WOW	UNIWARD	MIPOD	AVG
Pelee preact	0.6320	0.7201	0.9001	0.6445	0.7242
ResNet preact	0.6665	0.7656	0.8943	0.6686	0.7488

표 4.6 3계층 COVER, WOW, UNIWARD, MIPOD 분류

4.2.3 계층 분류 결과

표 4.7에서 N1 ~ N3는 4계층구조, N4, N5는 3계층구조의 결과이다. 표 4.8-(a)에서 4계층 구조는 PVD, LSB, UNIWARD를 순서대로 검출한 후 COVER와 WOW, MIPOD을 최종적으로 분류한다. Pelee 구조만을 사용했을 때 계층 분류 방법은 6개 클래스의 동시분류 방법보다 약 1.4%정도 높은 결과를 보인다. N2, N3에서 구조를 LSB를 검출하는 구조를 Resnet으로 변경하고 UNIWARD를 검출하는 구조를 Pelee_preact구조로 변경했을 때 각각 평균 80.28%, 80.54%로 분류 정확도가 상승했다.

3계층 구조는 LSB와 PVD를 분류하는 구조는 4계층 구조와 동일한 방법을 사용하고, 마지막 계층에서 COVER, UNIWARD, WOW, MIPOD을 분류한다. N4, N5에서 평균 정확도는 각각 81.52%, 79.29%의 정확도를 보인다. N4에서최하위 계층에 Resnet_preact구조를 사용했을 때 평균 정확도의 큰 상승을 보였다. 4계층에서 3계층으로 계층의 수를 줄였을 때 계층의 수에서 오는 정확도 하락이 감소하는 것을 볼 수 있었다. 하지만 상위 계층에서 LSB, PVD, UNIWARD를 분류하고 하위 계층에서 COVER, WOW MIPOD을 분류할 경우 기존의 동시분류보다 더 낮은 성능을 보였다.

Net	COVER	1LSB	PVD	WOW	UNIWARD	MIPOD	Average
6진	0.6810	0.8969	0.9983	0.7105	0.8653	0.5629	0.7858
N1	0.6910	0.9457	0.9995	0.6500	0.8885	0.6200	0.7991
N2	0.6908	0.9483	0.9995	0.6547	0.8996	0.6240	0.8028
N3	0.6935	0.9483	0.9995	0.6601	0.9021	0.6290	0.8054
N4	0.6657	0.9483	0.9995	0.7458	0.8740	0.6579	0.8152
N5	0.6525	0.9483	0.9995	0.6906	0.8719	0.5948	0.7929

(a)

Type	Net#1	Net#2	Net#3	Net#4
N1	Pelee	Pelee	Pelee	Pelee_pre
N2	Pelee	Resnet	Pelee	Pelee_pre
N3	Pelee	Resnet	Pelee_pre	Pelee_pre
N4	Pelee	Resnet	Resnet_pre	
N5	Pelee	Resnet	Pelee_pre	

(b)

표 4.7 계층적 CNN의 실험 결과 (a)분류 정확도 비교, (b)계층구조의 네트워크 구성

Ⅴ.결론

본 논문에서는 복수의 스테가노그래피 알고리즘을 분류하기 위해 계층적 CNN구조를 제안하였다. COVER, LSB, PVD, WOW, UNIWARD, MIPOD의 6개의 클래스를 분류하기 위해 계층적 CNN구조에서는 스테가노그래피 알고리즘들의 계층적 특징을 이용해 알고리즘들을 분류하는 방법을 사용했다. 계층적 구조는 동시분류에서 발생할 수 있는 클래스 간의 오분류를 줄이고, 각 알고리즘에 최적화된 CNN구조를 사용할 수 있는 장점이 있다. 동시분류 구조에서 COVER, LSB, PVD, WOW, UNIWARD, MIPOD를 분류했을 때 Pelee구조 기반의 방법이 약 78.58%로 가장 높은 성능을 보인다. 하지만 LSB 스테가노그래피 알고리즘이 2진 분류의 결과에서 평균 96%정도의 분류 정확도를 보이는데 비해, Pelee 구조 기반의 동시분류는 약 89% 정도로 LSB의 분류 정확도가 감소했다. 계층적 CNN구조에서 Pelee구조를 사용해 3계층구조를 구성했을 때 동시 분류방법이 89%의 분류 정확도를 보인 것에 비해 계층구조의 경우 약 94%로 LSB를 더 잘 분류한다. 또한 Pelee 단일 구조를 사용한 동시분류에 비해 각 클래스에 최적화된 구조를 사용한 계층구조의 경우 평균적인 분류정확도 또한 81.52%로 동시분류의 78.58%에 비해 더 높은 성능을 보인다.

복수의 스테가노그래피의 분류는 기존의 스테그아날리시스 분류에서 발생할 수 있는 문제점들을 개선하고, 검출된 스테가노그래피로부터 데이터를 추출하기 위한 첫 단계이다. 본 논문에서는 복수의 스테가노그래피 알고리즘을 분류하기 위한 계층적인 CNN 구조를 제안했다. 계층적 CNN구조는 기존의 동시 분류방법이 단일 네트워크에서 모든 알고리즘을 분류하는 것

과 달리 특정 알고리즘을 검출하고 다른 알고리즘들은 커버영상과 함께 다음 계층으로 보내어 계층적으로 데이터를 분류하는 방법을 사용했다. 제안 방법에서는 COVER, LSB, PVD, UNIWARD, WOW, MIPOD의 6개의 클래스에 대해서 데이터를 분류했는데, 동시 분류 방법보다 약 2%정도 향상된 성능을 얻었다. 제안 방법의 결과를 바탕으로 복수의 스테가노그래피를 분류하는데 있어서 단순히 CNN구조의 출력 클래스의 수를 증가시키는 것보다, 스테가노그래피 알고리즘의 특성을 분석하여 각 계층에서 특정한 스테가노그래피 알고리즘을 분류하는 계층적 분류 방법이 효과적임을 알 수 있다.

계층적 CNN 구조 기반의 스테그아날리시스는 기존의 스테그아날리시스의 커버, 스테고의 2진 분류가 아닌 계층적 네트워크를 통한 커버와 각 스테가노그래피 알고리즘들을 분류한다. 스테가노그래피 알고리즘의 분류는 데이터의 검출을 넘어 검출된 스테고 영상에 어떠한 스테가노그래피가 적용되었는지 분석하고, 스테고 영상으로부터 삽입된 데이터를 검출하기 위한 첫 단계이다. 제안 방법에서는 COVER, LSB, PVD, UNIWARD, WOW, MIPOD의 6개 클래스에 대해 계층적 분류에 대해 분석했고 계층적 구조를 통해 기존의 방법들보다 분류 성능에 있어서 향상된 성능을 보였다. CNN구조 기반의 스테그아날리시스는 스테가노그래피의 적용으로 인한 영상의 미소한 변화를 통해 스테고영상을 검출할 수 있어야 한다. 계층적 CNN구조는 스테가노그래피 데이터들의 특징을 바탕으로 기존의 단일구조 기반의 방법보다 더 최적화된 구조를 이용해 효과적으로 스테고영상을 검출할 수 있다.

Ⅵ.향후 계획

스테그아날리시스는 스테가노그래피가 적용된 데이터를 검출하는 방법이다. 스테가노그래피는 제 3자가 알지 못하게 사진, 동영상, 오디오 등 멀티미디에 매체에 데이터를 교묘하게 숨겨 통신한다. 스테가노그래피는 범죄 등에 같이 악용될 가능성이 있기 때문에 스테그아날리시스를 통한 스테가노그래피의 검출은 매우 중요하다.

최근 CNN기반의 스테그아날리시스는 이전의 수제특징 기반의 방법들에 비해 높은 검출 정확도를 보인다. 스테그아날리시스는 스테가노그래피가 적용된 데이터를 검출하지만 해당 매체로부터 데이터를 추출하는 일은 요원하다. 본 논문에서는 스테가노그래피로부터 데이터를 검출하기 위한 첫 번째 단계로 데이터에 어떠한 스테가노그래피가 적용되었는지를 분류하기 위해 COVER, 1LSB, WOW, UNIWARD, MIPOD 6개의 클래스를 분류했고 동시 분류하는 것보다 스테가노그래피의 특성을 이용해 네트워크를 계층적으로 구성하는 것이 더 효과적임을 보였다.

스테가노그래피 알고리즘은 영상에 어떠한 키를 적용하는가, 데이터의 삽입순서, 삽입용량 등에 따라 굉장히 다양한 방법을 사용할 수 있다. 이러한 다양한 스테가노그래피 알고리즘의 분류를 위해서는 본 논문에서 제안한 것과 같이 복수의 네트워크를 이용해 스테가노그래피 알고리즘들의 군을 나누어 분류하는 방법에 대한 연구가 필요할 것이다. 또한 분류하고자 하는 클래스의 수가 증가할수록 동시분류 방법과 계층적 분류 방법 모두에서 평균적인 분류 정확도가 감소하는 경향을 보인다. 이는 제안 방법에서 사용한 Pelee구조의 한계로 보인다. 따라서 클래스의 수에 증가에 따른 정

확도 감소를 상쇄하기 위한 강력한 구조에 대한 연구가 필요할 것이다.

참 고 문 헌

[1] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, “Digital image steganography: Survey and analysis of current methods,” Signal Processing, vol. 90, Issue 3, pp. 727-752, 2010.

[2] D. Neeta, K. Snehal and D. Jacobs, "Implementation of LSB Steganography and Its Evaluation for Various Bits," 2006 1st International Conference on Digital information management, Bangalore, pp. 173-178, 2007.

[3] 지선수, “LSB 기법 기반 최적의 이미지 스테가노그래피의 연구,” 한국산업정보학회논문지, 20(3), pp. 29-36, 2015.

[4] C. Chan and L. M. Cheng “Hiding data in images by simple LSB substitution,” Pattern Recognition, vol 37, pp. 469-474, 2004.

[5] K. Jung and K. Yoo, “Steganographic method based on interpolation and LSB substitution of digital images,” Multimedia Tools Applications, vol 74, pp. 2143 - 2155, 2015.

[6] G. Swain, “Adaptive and non-adaptive PVD steganography using overlapped pixel blocks,” Arab J Sci Eng, vol 43, pp. 7549 - 7562, 2018.

[7] K. Chang, P. S. Huang, T. Tu and C. Chang, "Adaptive image steganographic scheme based on Tri-way Pixel-Value Differencing," 2007 IEEE International Conference on Systems, Man and Cybernetics, Montreal, Que., pp. 1165-1170, 2007.

[8] S. Kang, H, Park. and J, Park, “Combining LSB embedding with modified Octa-PVD embedding,” Multimedia Tools Applications, vol. 79, pp. 21155 - 21175 2020.

[9] S. A. Thanekar and S. S. Pawar, "OCTA (STAR) PVD: A different approach of image steganopgraphy," 2013 IEEE International Conference on Computational Intelligence and Computing Research, Enathi, pp. 1-5, 2013.

[10] T. Pevny, T. Filler and P. Bas. “Using High-Dimensional image models to perform highly undetectable steganography,” Information Hiding, Calgary, Canada, 2010.

[11] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," 2012 IEEE International Workshop on Information Forensics and Security (WIFS), Tenerife, pp. 234-239, 2012.

[12] V. Holub, J. Fridrich and T. Denemark, “Universal distortion function for steganography in an Arbitrary Domain,” EURASIP Journal on Information Security, 2014.

[13] A. Ker, “Quantitative evaluation of pairs and RS steganalysis,” Proceedings Security, Steganography, and Watermarking of Multimedia Contents VI, vol. 5306, 2004.

[14] P. Pevny, P. Bas and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," in IEEE Transactions on Information Forensics and Security, vol. 5, no. 2, pp. 215-224, 2010.

[15] J. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” Neural Processing Letters 9, pp. 293 – 300 1999.
[16] J. Fridrich and J. Kodovsky, "Rich Models for Steganalysis of Digital Images," in IEEE Transactions on Information Forensics and Security, vol. 7, no. 3, pp. 868-882, 2012.
[17] A. Krizhevsky, H. Sutskever and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Advances in Neural Information Processing Systems 25(NIPS) 2012.
[18] J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, pp. 248-255, 2009.
[19] G. Xu, H. Wu and Y. Shi, "Structural Design of Convolutional Neural Networks for Steganalysis," in IEEE Signal Processing Letters, vol. 23, no. 5, pp. 708-712, 2016.
[20] S. Wu, S. Zhong and Y. Liu, “Deep residual learning for image steganalysis,” Multimedia Tools Application vol. 77, pp. 10437 – 10453, 2018.
[21] S. Kang, H. Park and J. Park, “CNN-Based Ternary Classification for Image Steganalysis,” Electronics, vol. 8, 2019.
[22] M. Naor and A. Shamir, “Visual cryptography,” Lecture Notes in Computer Science, vol. 950, Springer, 1994.
[23] Q. Huynh-Thu and M. Ghanbari, “Scope of validity of PSNR in image/video quality assessment,” Electronics letters vol. 44, Issue 13, pp. 800 – 801, 2008.
[24] V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R. P. Sheridan, and B. P. Feuston, “Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,” Journal of Chemical Information and Computer Sciences vol. 43(6), pp. 1947-1958, 2003.
[25] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," IEEE Transactions on Image Processing, vol. 11, no. 4, pp. 467-476, 2002.
[26] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.
[27] S. loffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Proceedings of the 32nd International Conference on Machine Learning, vol. 37, pp. 448-456, 2015.
[28] D. P. Mandic, "A generalized normalized gradient descent algorithm," IEEE Signal Processing Letters, vol. 11, no. 2, pp. 115-118, 2004.
[29] L. Bottou, “Stochastic Gradient Descent Tricks,” Neural Networks, Tricks of the Trade, Reloaded, Springer, vol. 7700, 2012.
[30] L. Mason, J. Baxter, P. Bartlett and M. Frean, “Boosting algorithms as gradient descent,” Proceedings of the 12th International Conference on Neural

Information Processing Systems, pp. 512 - 518, 1999.
[31] K. Chatfield, K. Simonyan, A. Vedaldi and A. Zisserman, “Return of the Devil in the Details: Delving Deep into Convolutional Nets,” Computer Vision and Pattern Recognition(CVPR), 2014.
[32] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.
[33] Y. Yuan, W. Lu, B. Feng, J. Weng, “Steganalysis with CNN Using Multi-channels Filtered Residuals,” Cloud Computing and Security Lecture Notes in Computer Science, vol. 10602, 2017.
[34] B. Li, W. Wei, A. Ferreira and S. Tan, "ReST-Net: Diverse Activation Modules and Parallel Subnets-Based CNN for Spatial Image Steganalysis," IEEE Signal Processing Letters, vol. 25, no. 5, pp. 650-654, 2018.
[35] X. Deng, B. Chen, W. Luo and D. Luo, “Fast and Effective Global Covariance Pooling Network for Image Steganalysis,” Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, pp. 230-234, 2019.
[36] K. He, X. Zhang, S. Ren and J. Sun. “Identity Mappings in Deep Residual Networks,” ECCV, Lecture Notes in Computer Science, vol 9908, 2016.
[37] S. Zagoruyko and N. Komodakis, “Wide Residual Networks,” Computer Vision and Pattern Recognition 2016.
[38] S. Xie, R. Girshick, P. Dollar, Z. Tu and K. He, “Aggregated Residual Transformations for Deep Neural Networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1492-1500, 2017.
[39] D. Han, J. Kim, J. Kim, “Deep Pyramidal Residual Networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 5927-5935, 2017.
[40] G. Huang, Z. Liu, L. Maaten and K. Q. Weinberger, “Densely Connected Convolutional Networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 4700-4708, 2017.
[41] R. J. Wang, X. LI and C. X. Ling, “Pelee: A Real-Time Object Detection System on Mobile Devices,” Advances in Neural Information Processing Systems 31(NIPS), 2018.
[42] V. Sedighi, R. Cogranne and J. Fridrich, "Content-Adaptive Steganography by Minimizing Statistical Detectability," IEEE Transactions on Information Forensics and Security, vol. 11, no. 2, pp. 221-234, 2016.
[43] http://pytorch.org