전처리 필터의 수가 CNN 기반 스테그아날리시스의 성능에 미치는 영향 분석

강상훈 [1], 박한훈 [1], 박종일 [2], 김산해 [3]
[1] 부경대학교, [2] 한양대학교, [3] 국방과학연구소
totohoon01@naver.com, hanhoon.park@pknu.ac.kr, jipark@hanyang.ac.kr

Analysis of the Effect of Number of Preprocessing Filters on the Performance of CNN-Based Steganalysis

Sanghoon Kang[1], Hanhoon Park[1], Jong-Il Park[2], Sanhae Kim[3]
[1]Pukyong National University, [2]Hanyang University, [3]Agency of Defense Development

요 약

본 논문에서는 CNN 기반 스테그아날리스 방법을 이용하여 입력 영상에 비밀 메시지가 삽입되었는지를 판별하고, 비밀 메시지가 삽입되었을 경우 WOW 와 UNIWARD 방법 중에 어떤 방법으로 삽입되었는지를 분류하고자 한다. 이를 위해 입력 영상으로부터 특징 정보를 추출하기 위해 사용되는 전처리(prepropcessing) 필터의 수가 분류 성능에 미치는 영향에 대해 분석한다. SRM 필터를 사용한 실험에서 필터의 수를 단순히 증가시키는 것은 성능 향상이 도움이 되지 않으며, 효과적인 필터를 선별해서 사용하는 것이 보다 우수한 성능을 가짐을 확인하였다.

1. 서론

영상 스테그아날리시스(steganalysis)는 영상 스테가노그래피(steganography)를 이용해 숨긴 정보를 찾는 기법을 의미한다. 최근의 스테그아날리시스의 경향은 CNN(convolutional neural network)을 활용하여 특징 정보를 추출하고 분류하는 방법을 사용한다. CNN 의 사용 유무와 상관없이 스테그아날리시스는 기본적으로 입력 영상에 고주파 필터(HPF)를 적용하여 특징 정보 추출을 하는데, 다수의 필터를 사용할수록 좋은 성능을 보이는 경향이 있다.

CNN 을 사용하지 않고 미리 설계된 필터에 의해서만 특징 정보를 추출하는 대표적인 방법인 SRM[1]은 30 개의 필터를 사용해 영상으로부터 많은 특징을 추출하고자 했다. CNN 기반 스테그아날리시스에서는 Multi-channel CNN[2]은 Xu 와 Wu 에 의해 제안된 CNN 기반 스테그아날리시스의 방법[3]에서 필터의 수를 증가시켜 더 좋은 결과를 얻었고 ReST-Net[4]은 Gabor 필터와 SRM 필터를 사용해 많은 특징을 얻고자 했다. 하지만 단순히 많은 수의 필터를 사용해 성능을 높이는 것에는 한계가 있다.

본 논문에서는 SRM 필터 중 특징을 잘 추출하는 10 개의 필터만을 사용해 30 개의 필터를 모두 사용했을 때보다 더 좋은 성능을 보이는 CNN 구조와 결과를 통해 구조에 따라서 적절한 필터의 수를 사용할 필요가 있음을 보이고자 한다.

2. 관련 연구

SRM 은 30 개의 선형, 비선형 필터를 입력 영상에 적용해

많은 특징 맵(feature map)을 생성함으로써, 입력 영상으로부터 가능한 많은 정보를 추출하기 위한 목적을 지녔다(Fig. 1 참조). CNN 기반 스테그아날리시스는 SRM 과 같은 기존의 방법들과 마찬가지로 영상에 고주파 필터를 적용하여 CNN 구조를 학습하고 분류하는 과정을 거친다. CNN 기반 스테그아날리시스는 다양한 필터들을 사용하는데, Multi-channel CNN 과 Rest-Net 의 경우 각각 3 개, 44 개의 필터를 사용했다(Fig. 2 참조).

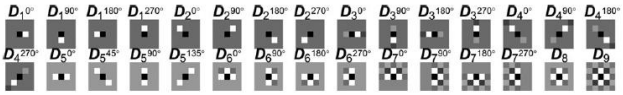


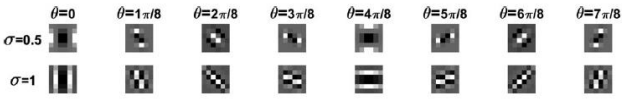
Fig. 1. 30 개의 SRM 필터. 방향에 따라 다양한 특징 정보를 추출한다.

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

$$K_1 = \frac{1}{4}\begin{pmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{pmatrix}$$

$$K_2 = \frac{1}{4}\begin{pmatrix} 2 & -1 & 2 \\ -1 & -4 & -1 \\ 2 & -1 & 2 \end{pmatrix}$$

(a) Multi-channel CNN의 필터



(b) ReST-Net에서 사용된 Gabor 필터

Fig. 2. Multi-channel CNN 과 ReST-Net 에서 사용된 필터. ReST-Net 은 Fig. 1의 SRM 필터도 함께 사용한다.

3. 제안 방법

제안 방법은 커버(cover) 영상과 WOW[5], UNIWARD[6]를 적용한 스테고(stego) 영상을 분류하기 위한 CNN 기반 3 진 분류기를 설계했다(Fig. 3 참조). 설계된 분류기에 대해 고주파 필터의 수의 영향을 분석하기 위해 기존 SRM 필터 30 개와 이로부터 선별된 10 개의 필터를 사용한 분류 결과를 비교, 분석했다.

각각 30 개와 10 개의 필터를 사용하는 것 이외 모든 파라미터는 동일하게 설정된다. 고주파 필터를 통과한 256x256 의 영상은 CONV1 계층에서 60 개의 특징 맵을 생성한다. 각 특징 맵은 계층마다 2 배씩 증가하여 1920 개의 1x1 특징 맵이 된다.

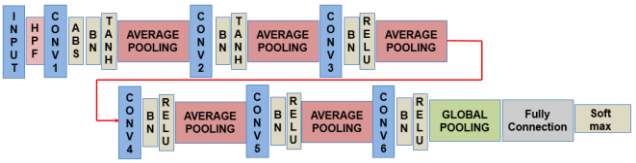


Fig. 3. 제안된 3 진 분류기.

SRM 필터로부터 분류에 효과적인 필터를 선별하는 과정은 다음과 같다. 커버 영상과 스테고 영상에 대해 SRM 의 각 필터를 적용한 후 필터링된 두 영상 사이의 차를 구한다. 모든 필터에 대한 차의 평균을 구한 후, 평균보다 큰 차를 보이는 필터를 찾는다(Fig. 4 참조).

Fig. 4. SRM 필터로부터 분류 성능이 높은 필터 선별 과정. 각 필터를 적용한 후 커버 영상과 스테고 영상 사이의 차는 필터에 따라 큰 차이(예, 1.229 와 7.234)를 보인다.

3. 실험 결과

256x256 의 200,000 개의 학습(training) 영상과 5,000 개의 테스트(test) 영상을 사용했다. Optimizer 로 Momentum optimizer 을 사용했고 momentum 의 값은 0.9 를 사용한다. Learning rate 는 0.001 에서 시작해서 매 5,000 번마다 90%로 감소한다. 표 1 은 30 개의 SRM 필터와 10 개의 선별된 SRM 필터에 따른 커버 영상, WOW, UNIWARD 에 의한 스테고 영상의 분류 결과를 정확도로 계산한 것이다.

30 개의 SRM 필터를 사용했을 때 약 66%의 분류 정확도를 보인다. 하지만 본 논문에서 제안한 방법대로 선별된 분류 성능이 우수한 10 개의 필터만 사용했을 때는 분류 정확도가 약 70%로 4%정도 더 높은 정확도를 보였다. 즉, 적은 수의 필터를 이용해 학습한 것이 더 정확한 분류를 하는 것을 알 수 있다. 따라서 CNN 기반 스테그아날리시스에서 필터의 수를 단순히 증가시키는 것은 성능 향상이 도움이 되지 않으며, 효과적인 필터를 선별해서 사용하는 것이 보다 우수한 성능을 가짐을 확인할 수 있다.

표 1. 필터 수에 따른 분류 정확도(인식율[%]) 결과.

필터의 개수	Total	Cover	WOW	UNIWARD
30	66.262	75.485	51.845	71.455
10	70.098	78.165	72.690	59.440

4. 결론

본 논문에서는 CNN 기반 스테그아날리시스에서 전처리 필터의 수에 따른 성능 변화를 SRM 필터를 이용하여 분석했다. 커버 영상과 WOW, UNIWARD 를 이용하여 생성된 스테고 영상을 분류하는 실험에서 30 개의 SRM 필터를 사용했을 때보다 30 개의 필터 중 특징 추출 성능이 높은 10 개의 필터를 사용했을 때 분류 정확도에 있어서 더 높은 결과를 보였다. 이를 통해 CNN 기반 스테그아날리시스에서 특징 맵의 수를 증가시키기 위해 단순히 필터의 수를 증가시키는 것은 한계가 있으며 특징 추출 성능이 우수한 필터를 선별해서 사용하는 것이 더 효과적임을 알 수 있었다.

감사의 글

본 연구는 방위사업청 및 국방과학연구소의 재원에 의해 설립된 신호정보 특화연구센터 사업의 지원을 받아 수행된 것임.

참고 문헌

[1] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," IEEE Trans. Inf. Forensics Security, vol. 7, no. 3, pp. 868-882, Jun. 2012.

[2] Y. Yuan, W Lu, B, Feng and J. Weng "Steganalysis with CNN using multi-channels filtered residuals", ICCCS 2017, LNCS 10602, pp. 110-120, 2017.

[3] G. Xu and H. Wu, "Structure design of convolution neural networks for steganalysis", IEEE Signal Processing Letters, vol. 23, no. 5, pp. 708-712, 2016.

[4] 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.

[5] V. Holub, J. Fridrich, "Designing steganographic distortion using directional filters," IEEE Workshop on Information Forensic and Security, 2012.

[6] V. Holub, J. Fridrich, T. Denemark, "Universal distortion function for steganography in an arbitrary domain," EURASIP Journal on Information Security, 2014.