SRM significantly reduces the P_E on detecting ADS, but suffers from greatly increase of P_E on detecting the original WOW.

Table 1: Testing error of WOW and ADS-WOW under different steganalyzers with a relative payload of 0.4 bpp

Error Rate(%)	WOW			ADS-WOW		
Steganalyzer	P_{MD}	P_{FA}	P_E	P_{MD}	P_{FA}	P_E
Xu's CNN	25.54	26.60	26.07	86.71	26.60	56.66
Ye's CNN	21.34	18.55	19.95	74.26	18.55	46.41
Wu's CNN	28.04	35.17	31.61	75.34	35.17	55.26
Multi-CNN	7.01	58.43	32.72	56.28	58.43	57.36
SRM	25.23	25.77	25.50	47.81	25.77	36.79
SRM (retrained)	23.90	44.72	34.31	13.33	44.72	29.03

4 CONCLUSIONS

In this paper, we propose a method of iteratively constructing robust enhanced cover images that can resist the neural networks for steganalysis and the intensity of adversarial noise is controllable. The stegos, obtained by using the constructed images as cover, can effectively avoid the detection of network-based steganalyzers. Besides, we also consider how to simultaneously fight against network-based steganalyzers and SRM+EC and define the comprehensive security criterion P_E^B under the two systems. We have made a tradeoff between the two systems and evaluated the performance of our model using the BOSSbase dataset, the WOW steganography method and three state-of-the-art networks. Results show the effectiveness of our method and comprehensive security level has been improved.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China under Grant U1636201 and 61572452. The authors would like to thank DDE Laboratory of SUNY Binghamton for sharing the source code of steganography, steganalysis and ensemble classifier on the webpage (http://dde.binghamton.edu/download/).

REFERENCES

[1] Shumeet Baluja and Ian Fischer. 2017. Adversarial Transformation Networks: Learning to Generate Adversarial Examples. *arXiv preprint arXiv:1703.09387* (2017).

[2] Patrick Bas, Tomáš Filler, and Tomáš Pevnỳ. 2011. âĂİ Break Our Steganographic SystemâĂİ: The Ins and Outs of Organizing BOSS. In *Information Hiding*. Springer, 59–70.

[3] Tomas Denemark, Vahid Sedighi, Vojtech Holub, Rémi Cogranne, and Jessica Fridrich. 2014. Selection-channel-aware rich model for steganalysis of digital images. In *Information Forensics and Security (WIFS), 2014 IEEE International Workshop on*. IEEE, 48–53.

[4] Jessica Fridrich and Jan Kodovsky. 2012. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security* 7, 3 (2012), 868–882.

[5] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

[6] Linjie Guo, Jiangqun Ni, Wenkang Su, Chengpei Tang, and Yun-Qing Shi. 2015. Using statistical image model for JPEG steganography: uniform embedding revisited. *IEEE Transactions on Information Forensics and Security* 10, 12 (2015), 2669–2680.

[7] Vojtech Holub and Jessica Fridrich. 2012. Designing steganographic distortion using directional filters. In *Information Forensics and Security (WIFS), 2012 IEEE International Workshop on*. IEEE, 234–239.

[8] Vojtěch Holub, Jessica Fridrich, and Tomáš Denemark. 2014. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security* 2014, 1 (2014), 1.

[9] Sandy Huang, Nicolas Papernot, Ian Goodfellow, Yan Duan, and Pieter Abbeel. 2017. Adversarial attacks on neural network policies. *arXiv preprint arXiv:1702.02284* (2017).

[10] Jan Kodovskỳ and Jessica Fridrich. 2012. Steganalysis of JPEG images using rich models. In *Media Watermarking, Security, and Forensics 2012*, Vol. 8303. International Society for Optics and Photonics, 83030A.

[11] Jan Kodovsky, Jessica Fridrich, and Vojtěch Holub. 2012. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security* 7, 2 (2012), 432–444.

[12] Jernej Kos, Ian Fischer, and Dawn Song. 2017. Adversarial examples for generative models. *arXiv preprint arXiv:1702.06832* (2017).

[13] Sarra Kouider, Marc Chaumont, and William Puech. 2013. Adaptive steganography by oracle (ASO). In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 1–6.

[14] Bin Li, Ming Wang, Jiwu Huang, and Xiaolong Li. 2014. A new cost function for spatial image steganography. In *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 4206–4210.

[15] Jiajun Lu, Theerasit Issaranon, and David Forsyth. 2017. Safetynet: Detecting and rejecting adversarial examples robustly. *arXiv preprint arXiv:1704.00103* (2017).

[16] Tomáš Pevnỳ, Tomáš Filler, and Patrick Bas. 2010. Using high-dimensional image models to perform highly undetectable steganography. In *International Workshop on Information Hiding*. Springer, 161–177.

[17] Yinlong Qian, Jing Dong, Wei Wang, and Tieniu Tan. 2015. Deep learning for steganalysis via convolutional neural networks. *Media Watermarking, Security, and Forensics* 9409 (2015), 94090J–94090J.

[18] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. 2013. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199* (2013).

[19] Weixuan Tang, Haodong Li, Weiqi Luo, and Jiwu Huang. 2014. Adaptive steganalysis against WOW embedding algorithm. In *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*. ACM, 91–96.

[20] Songtao Wu, Shenghua Zhong, and Yan Liu. 2017. Deep residual learning for image steganalysis. *Multimedia Tools and Applications* (2017), 1–17.

[21] Songtao Wu, Sheng-Hua Zhong, and Yan Liu. 2016. Steganalysis via Deep Residual Network. In *Parallel and Distributed Systems (ICPADS), 2016 IEEE 22nd International Conference on*. IEEE, 1233–1236.

[22] Songtao Wu, Sheng-hua Zhong, and Yan Liu. 2017. Residual convolution network based steganalysis with adaptive content suppression. In *Multimedia and Expo (ICME), 2017 IEEE International Conference on*. IEEE, 241–246.

[23] Guanshuo Xu, Han-Zhou Wu, and Yun-Qing Shi. 2016. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters* 23, 5 (2016), 708–712.

[24] Jian Ye, Jiangqun Ni, and Yang Yi. 2017. Deep learning hierarchical representations for image steganalysis. *IEEE Transactions on Information Forensics and Security* 12, 11 (2017), 2545–2557.