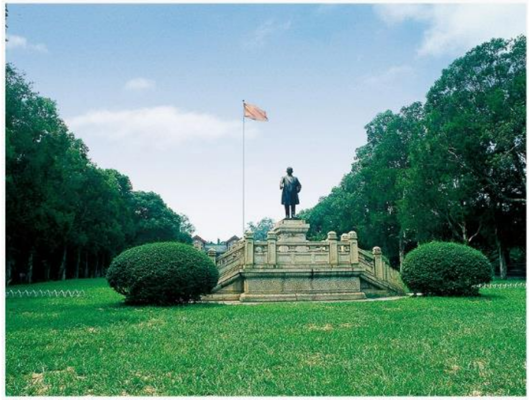
APPROACHING OPTIMAL EMBEDDING IN AUDIO STEGANOGRAPHY WITH GAN

Jianhua Yang[1], Huilin Zheng[1], Xiangui Kang[1*], Yun-Qing Shi[2]
1Guangdong Key Lab of Information Security,
Sun Yat-Sen University, Guangzhou, China
[2]Department of ECE, New Jersey Institute of Technology, Newark, NJ, USA



Acknowledgement: This work was supported by NSFC (Grant nos. U1536204, 61772571).

Outline

中山大學 SUN YAT-SEN UNIVERSITY

INTRODUCTION

Introduction

- Steganography
  - a kind of covert communication method which uses human perceptual redundancy to hidden messages into digital media, such as images, audio or video, without raising any suspicion.
- Challenges
  - Deep learning based audio steganalysis
- Hand-crafted methods
  - Can not adjust the embedding cost automatically according to the deep learning based steganalyzers.

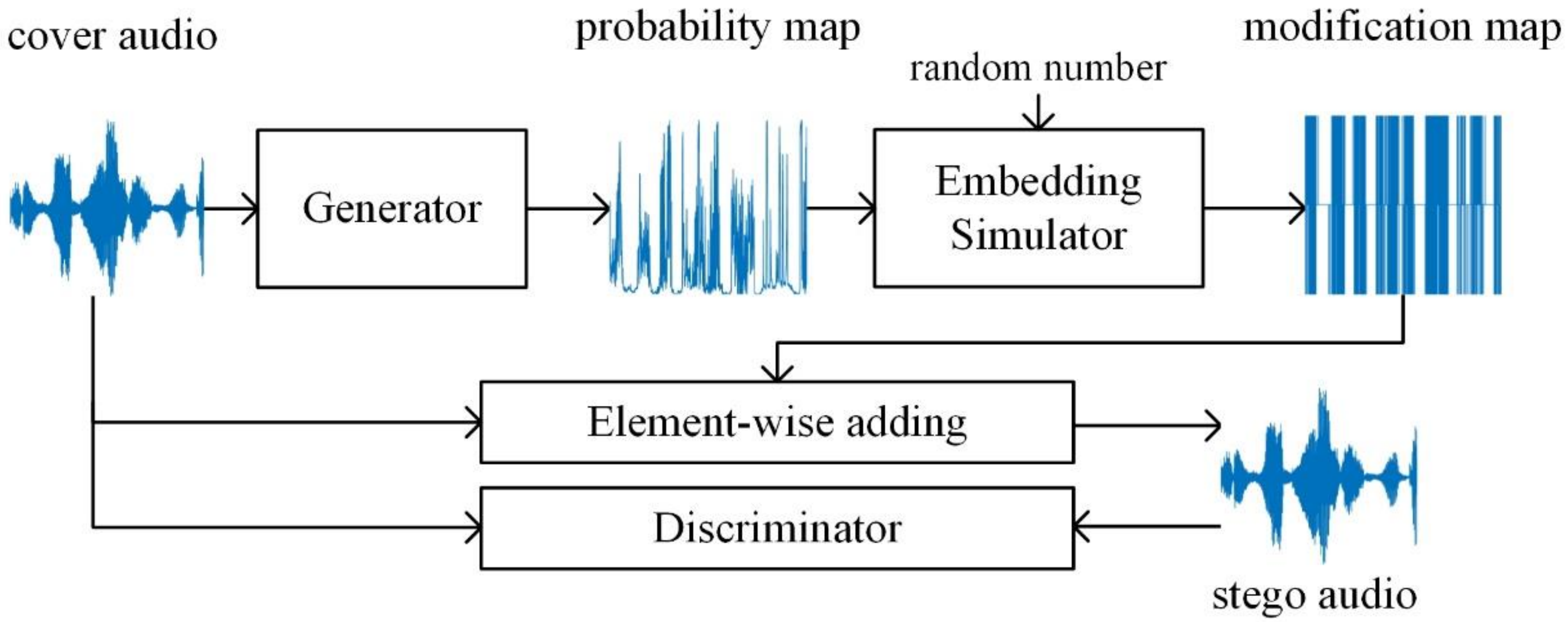
Introduction

- Image steganography based on Generative Adversarial Networks
    - ASDL-GAN, UT-GAN, and JS-GAN.
- GAN-based audio steganography
    - “probability map generation” approach
    - embedding for temporal domain

THE PROPOSED FRAMEWORK

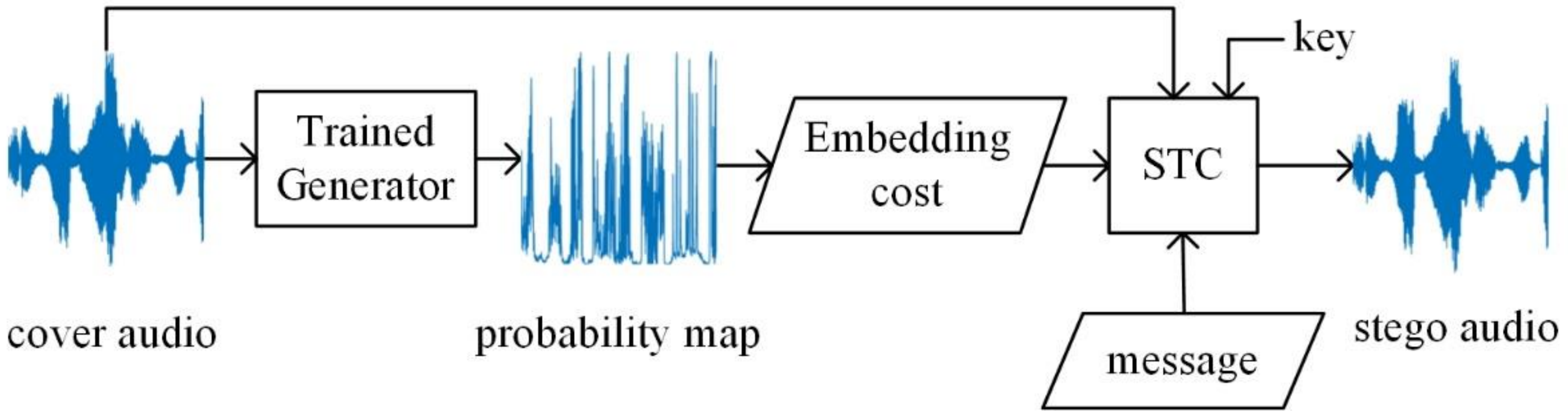
Two phases

1. Training phase: training the framework to obtain generator for "probability map generation"



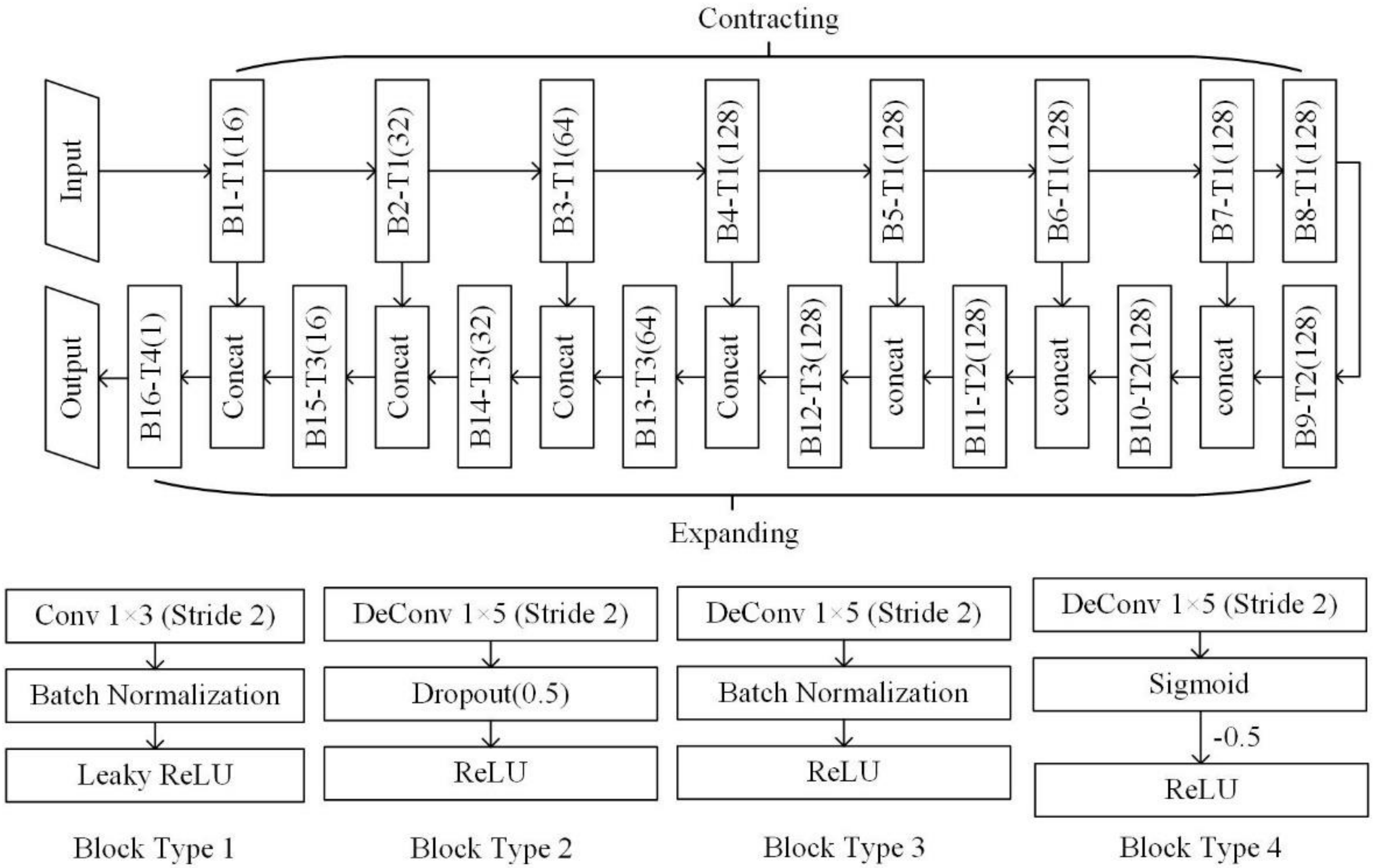
Two phases

2. Steganography: using the generator for practical applications of steganography with STC



Generator

- The U-Net based generator is used to generate an embedding probability for each sample of the cover audio.
- Four types of blocks



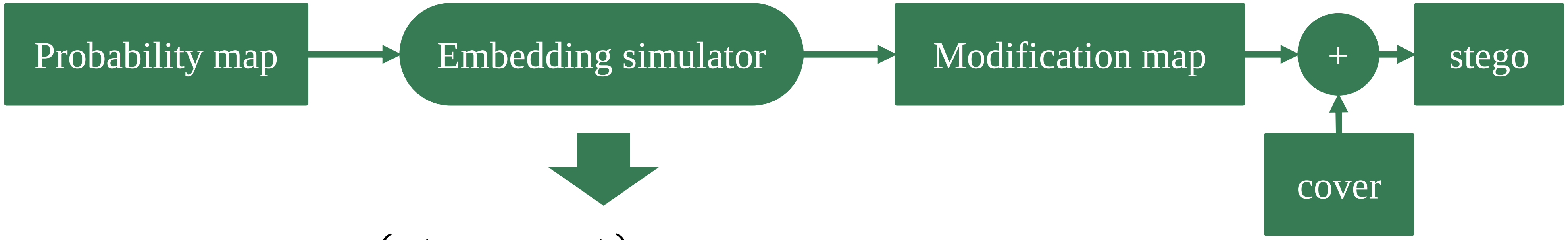
Embedding Simulator

- Embedding simulator is used to translate the probability map into modification map in training phase.
- In conventional steganography methods, the optimal embedding simulator which can be used to convert the probability to modification, is a three-stage staircase function and cannot backpropagate gradients through neural network.

$$m_i = \begin{cases} -1, & if\ r_i < \frac{p_i}{2} \\ 1, if\ r_i > 1 - \frac{p_i}{2} & \\ 0, & otherwise \end{cases}$$

Embedding Simulator

❑ Double-tanh function



$$m_i = -0.5 \times \tanh\big(\lambda(p_i - 2 \times r_i)\big) + 0.5 \times \tanh(\lambda(p_i - 2 \times (1 - r_i)))$$

$\{m_i\} \in [-1,1]^{1\times n}$: modification map

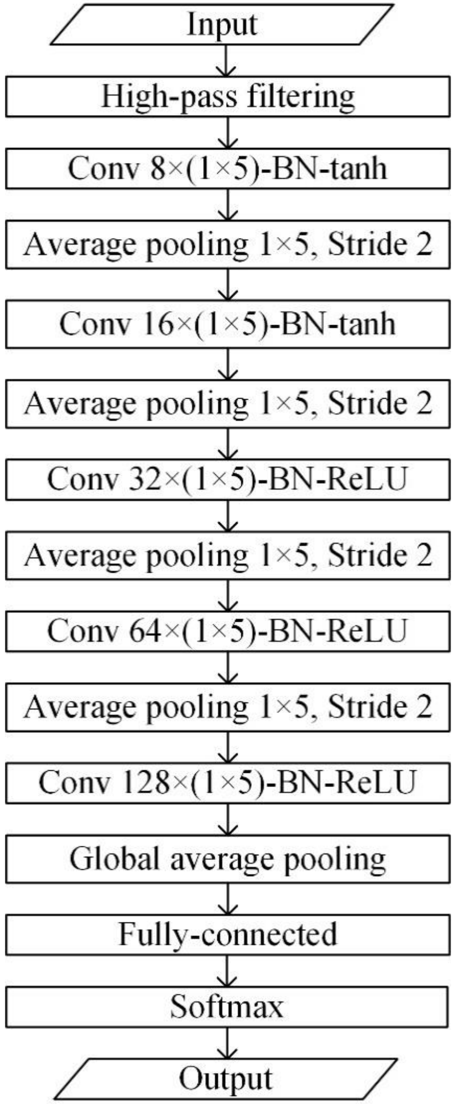
$\{p_i\} \in [0,0.5]^{1\times n}$: probability map

$\{r_i\} \in [0,1]^{1\times n}$: random numbers obeying uniform distribution ranging from 0 to 1

λ =60

Discriminator

- The discriminator is composed of
  - a high-pass filtering (HPF) layer
  - five convolutional blocks
  - average pooling layers (one global average pooling)
  - a fully-connected layer
  - a softmax layer



Loss Function

SUN YAT-SEN UNIVERSITY

- The discriminator loss function: $l_D = -\sum_{i=1}^{2} y_i \log(y_i')$
- The generator loss function

$$l_G = \alpha \times l_G^1 + \beta \times l_G^2$$
$$l_G^1 = -l_D$$
$$l_G^2 = (capacity - n \times payload),$$

Where

$$capacity = \sum_{i=1}^{n} (-p_i^{+1} \log_2 p_i^{+1} - p_i^{-1} log_2 p_i^{-1} - p_i^0 \log_2 p_i^0)$$

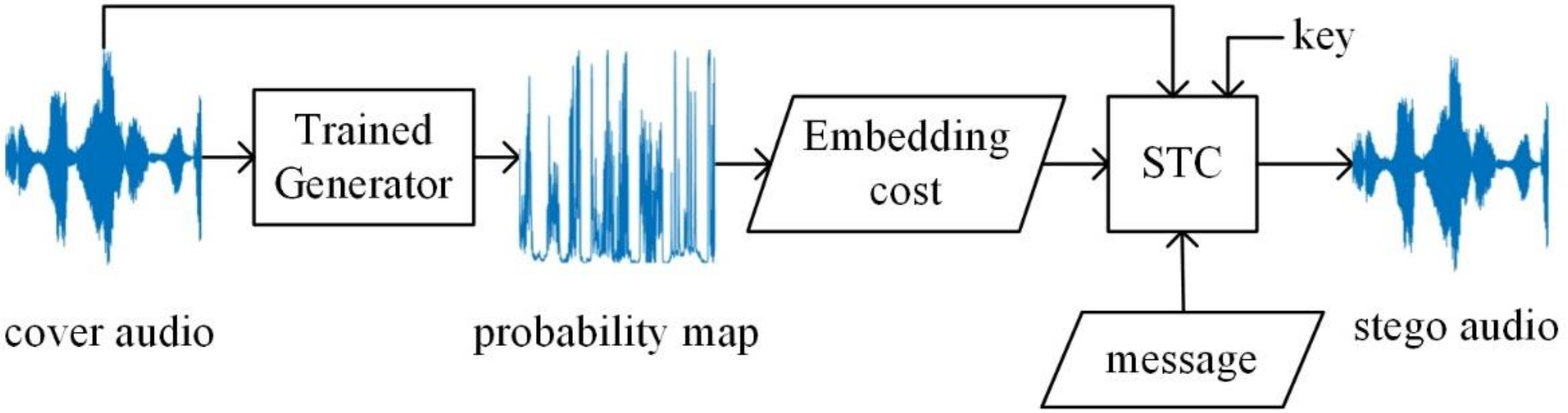
$$p_i^{+1} = p_i^{-1} = \frac{p_i}{2}, \qquad p_i^{+1} + p_i^{-1} + p_i^0 = 1$$

probability of modification value to be +1

probability of modification value to be -1

Embedding

- Cost calculation
  - $\rho_i = \ln(\frac{2}{p_i} - 2)$
- Embedding message
  - stego = STC(cover, Message, ρ)



EXPERIMENTAL RESULTS

Datasets and Settings

- Dataset
  - UME-ERJ: sampling rate is 16 kHz, 20,000 speech clips with length of 1 second
  - WSJ0: sampling rate is 16 kHz, 4,000 speech clips from original testing set and 30,000 from original training set with length of 1 second
- Usage
  - UME – used to train the proposed framework.
  - WSJ – steganography dataset, used to evaluate the security of different steganography

Datasets and Settings

- Hyperparamters
  - Learning rate: 0.001 for 0.4bps(bit per sample), 0.0001 for other embedding rate
    - Finetune: 0.4bps→0.3bps & 0.5bps, 0.3bps → 0.2bps, etc.
  - Batch size: 64
  - Training iterations: 7,000
  - Adam optimizer
  - Weights of the generator loss function: α=1, β=10^{-7}
- Steganalysis method: ChenNet[1], a CNN based audio steganalysis

[1] B. Chen, W. Luo, and H. Li, "Audio steganalysis with convolutional neural network," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec 2017, Philadelphia, PA, USA, June 20-22, 2017*, pp. 85–90.

Datasets and Settings

Selection of β

- Fixed $\alpha = 1$, then selected β from $\{10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}\}$
- When β was less than 10^{-7}, the capacity calculated by

$$capacity = \sum_{i=1}^{n} (-p_i^{+1} \log_2 p_i^{+1} - p_i^{-1} log_2 p_i^{-1} - p_i^{0} \log_2 p_i^{0})$$

  cannot be well fitted to the desired embedding capacity
- The security decreased as β increased from 10^{-7}

Table 1. Detection error rate (%) of different value of β using CNN based steganalyzer

$\boldsymbol{\beta}$	$\mathbf{10^{-7}}$	$\mathbf{10^{-6}}$	$\mathbf{10^{-5}}$	$\mathbf{10^{-4}}$
detection error	38.24	35.50	32.28	29.24

Adversarial Training

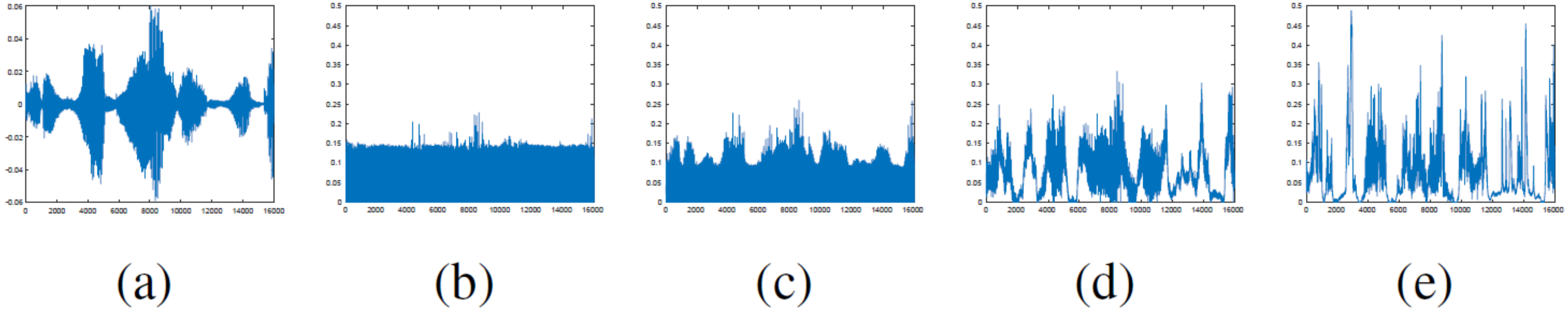


Fig. 1. Simulating results for the proposed framework with different training iterations when embedding rate is 0.4 bps for `00aa010a.wav' in WSJ. (a) is the origin audio, and (b)-(e) are the embedding probability generated by GAN trained after 500, 1,000, 2,000 and 7,000 iterations respectively.

Table 2. Detection error rate with respect to different training interations(%) when embedding rate is 0.4 bps.

Iteration	500	1000	2000	3000	4000	5000	6000	7000
detection error	28.16	33.35	37.69	36.57	38.56	37.88	37.50	38.82

Comparison with Existing Methods

Additional experiment

Dataset for GAN training	Dataset for embedding
UME	WSJ
WSJ	UME

Comparison methods

- LSB Matching [1]
- AAC based audio steganography [2]

[1] T. Sharp, “An implementation of key-based digital signal steganography,” in Information Hiding, 4th International Workshop, IHW 2001, Pittsburgh, PA, USA, April 25-27, 2001, Proceedings, pp. 13–26.
[2]W. Luo, Y. Zhang, and H. Li, “Adaptive audio steganography based on Advanced Audio Coding and Syndrome-Trellis Coding,” in Digital Forensics and Watermarking - 16th International Workshop, IWDW 2017, Magdeburg, Germany, August 23-25, 2017, Proceedings, pp. 177–186.

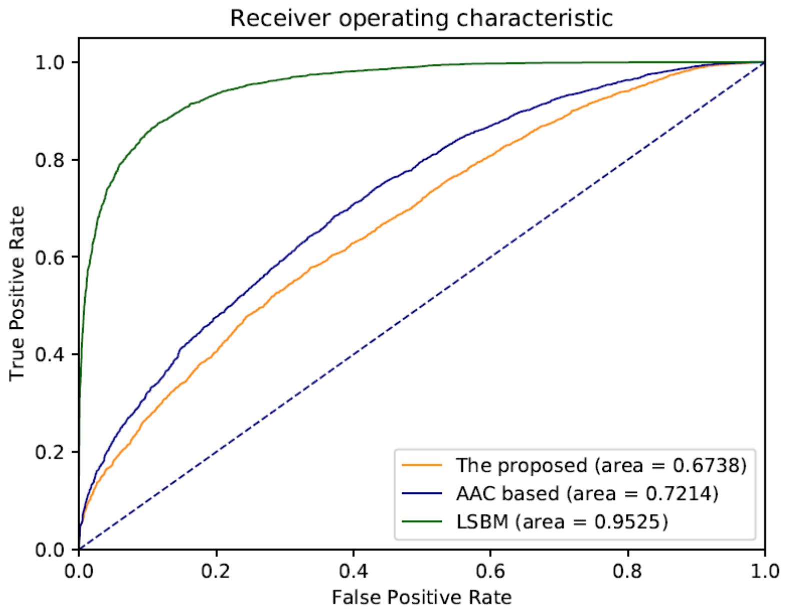
Comparison with Existing Methods

中山大學 SUN YAT-SEN UNIVERSITY

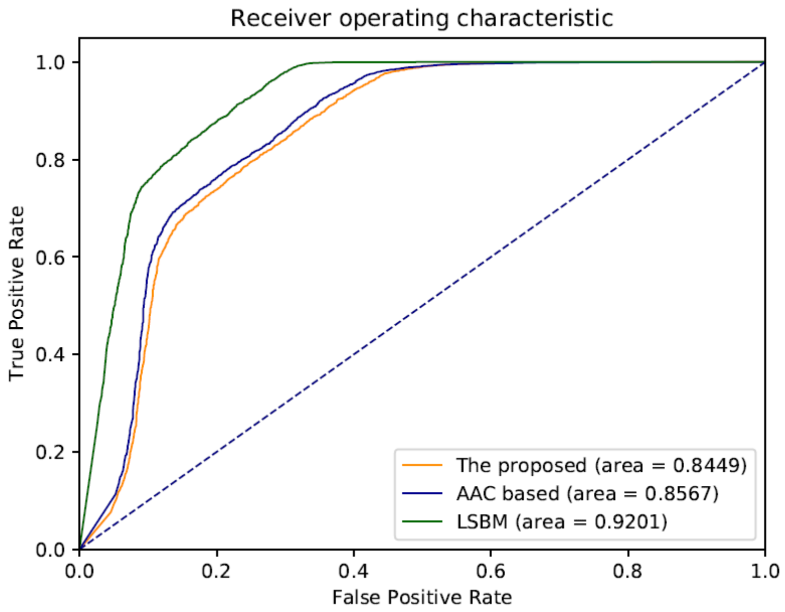
Table 3. Detection error rate (%) of different steganography.

Training dataset for proposed framework	steganography	Embedding rate (bps)				
		0.1	0.2	0.3	0.4	0.5
UME	LSB Matching	37.76	25.29	16.83	12.74	8.51
	AAC based	47.68	43.92	38.55	34.71	30.16
	The proposed	**48.34**	**45.10**	**41.95**	**38.24**	**33.26**
WSJ	LSB Matching	24.45	18.39	17.15	16.12	15.78
	AAC based	37.89	29.52	24.42	22.13	20.40
	The proposed	**40.93**	**31.86**	**26.61**	**23.01**	**21.07**

Comparison with Existing Methods



(a) training GAN with UME



(b) training GAN with WSJ

CONCLUSION

Conclusion

SUN YAT-SEN UNIVERSITY

- In this work, we have proposed a framework to learn the embedding probability automatically for audio steganography.
- The experimental results showed that the proposed framework can learn the adaptive embedding probability automatically and obtain better security than hand-crafted audio steganography LSB matching and AAC based method.
- In future research, we will investigate automatic cost learning for audio steganography in the frequency domain and coded domain.

Thank you!