

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN ĐIỆN TỬ - VIỄN THÔNG



ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC

# XÂY DỰNG HỆ THỐNG PHÁT HIỆN ÂM THANH BẤT THƯỜNG SỬ DỤNG AUTOENCODERS

Sinh viên thực hiện: Nguyễn Minh Hiếu  
Lớp Điện tử 03 – K60

Giảng viên hướng dẫn: TS. Nguyễn Thị Kim Thoa  
TS. Hàn Huy Dũng

Hà Nội, tháng 7/2020

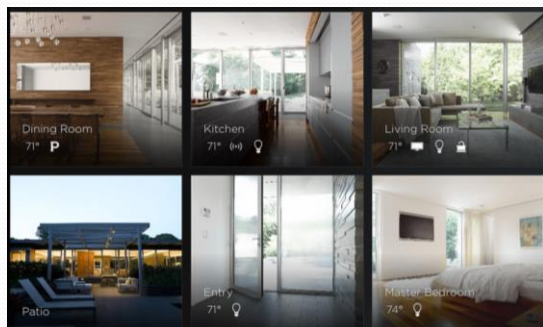
## Nội dung



- 1 Giới thiệu chung
- 2 Mô hình hệ thống
- 3 Các mô hình mạng neural được sử dụng
- 4 Thí nghiệm và kết quả
- 5 Kết luận

# Giới thiệu chung

## Nhu cầu về giám sát hiện nay



**Giám sát nhà ở**



**Giám sát nhà máy**

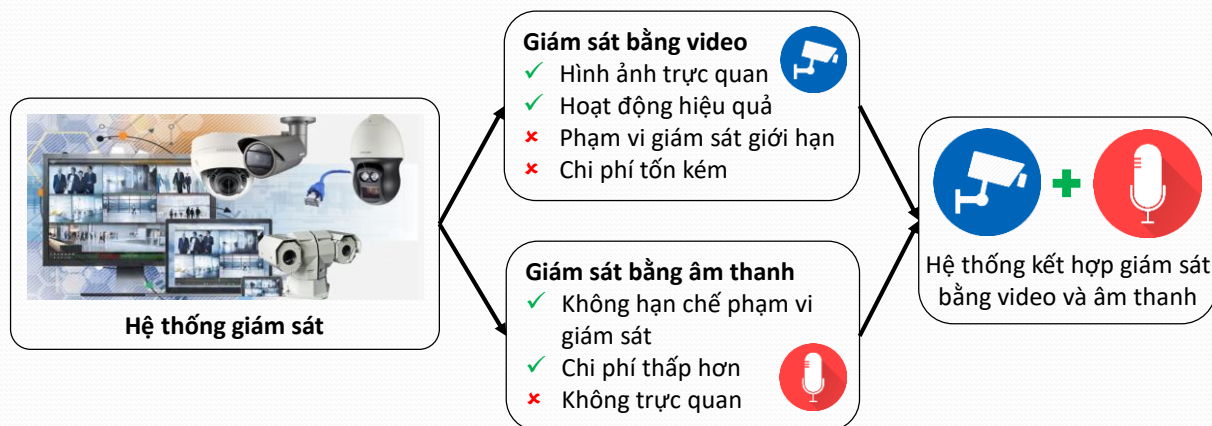
- Giám sát giúp theo dõi hoạt động, đảm bảo an ninh, cảnh báo nguy hiểm
- Các hệ thống giám sát được sử dụng ở nhiều nơi

15/07/2020

Signal Processing and Radio Communications Lab

3

## Tổng quan hệ thống giám sát



**➡➡➡ Hệ thống phát hiện âm thanh bất thường**

15/07/2020

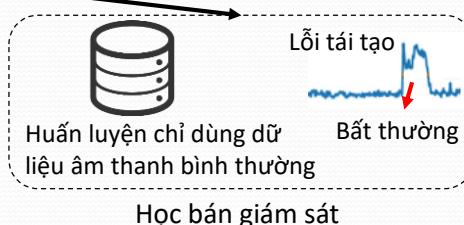
Signal Processing and Radio Communications Lab

4

## Hệ thống phát hiện âm thanh bất thường



Phương pháp phát hiện âm thanh bất thường sử dụng mạng neural



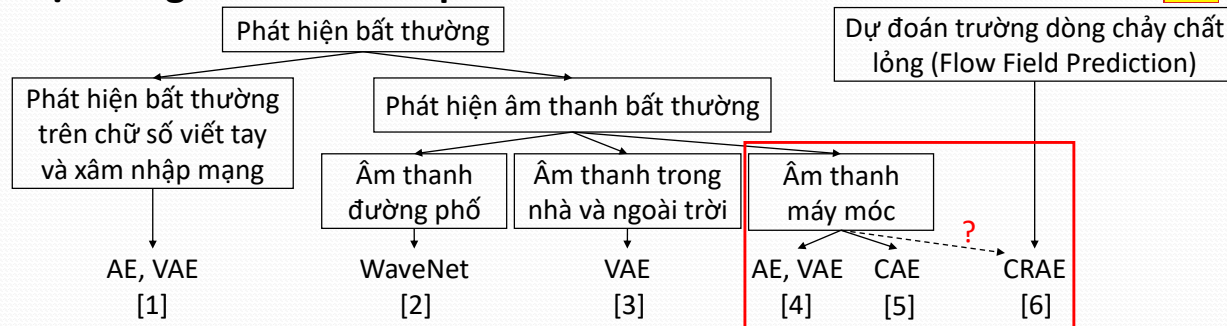
Autoencoder  
và các biến thể

15/07/2020

Signal Processing and Radio Communications Lab

5

## Một số nghiên cứu liên quan



- AE: Autoencoder
- VAE: Variational Autoencoder
- CAE: Convolutional Autoencoder
- CRAE: Convolutional Recurrent Autoencoder

[1] Jinwon An and Sungzoon Cho, "Variational Autoencoder based Anomaly Detection using Reconstruction Probability," in *2015-2 Special Lecture on IE*, December 2015

[2] T. Hayashi, T. Komatsu, R. Kondo, T. Toda and K. Takeda, "Anomalous Sound Event Detection Based on WaveNet," in *2018 26th European Signal Processing Conference*, Rome, Italy, September 2018

[3] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi and N. Harada, "Unsupervised Detection of Anomalous Sound Based on Deep Learning and the Neyman-Pearson Lemma," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 212-224, January 2019

[4] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo and Y. Kawaguchi, "Anomalous Sound Detection Based on Interpolation Deep Neural Network," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, Barcelona, Spain, May 2020

[5] R. Müller, F. Ritz, S. Illium and C. Linhoff-Popien, "Acoustic Anomaly Detection for Machine Sounds based on Image Transfer Learning," in *arXiv:2006.03429 [eess.AS]*, June 2020

[6] S. R. Bukka, A. R. Magee and R. K. Jaiman, "Deep Convolutional Recurrent Autoencoders for Flow Field Prediction," *arXiv:2003.12147 [physics.flu-dyn]*, March 2020

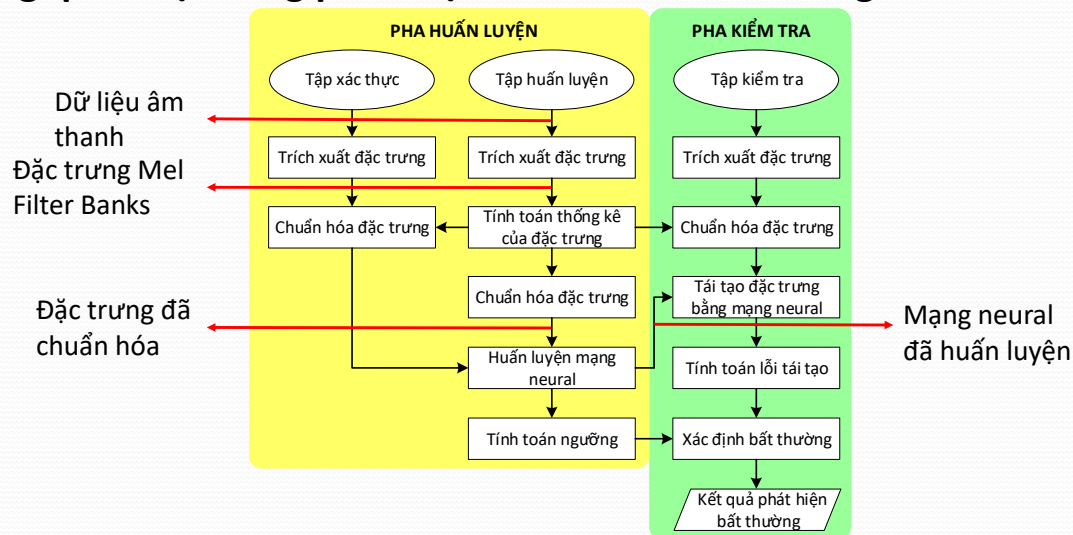
15/07/2020

Signal Processing and Radio Communications Lab

6

# Mô hình hệ thống

## Tổng quan hệ thống phát hiện âm thanh bất thường

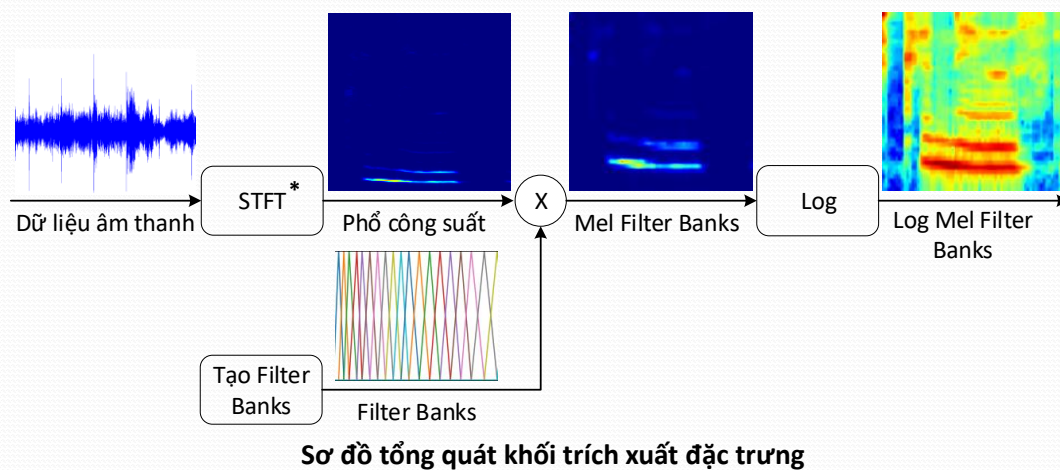


15/07/2020

Signal Processing and Radio Communications Lab

7

## Trích xuất đặc trưng

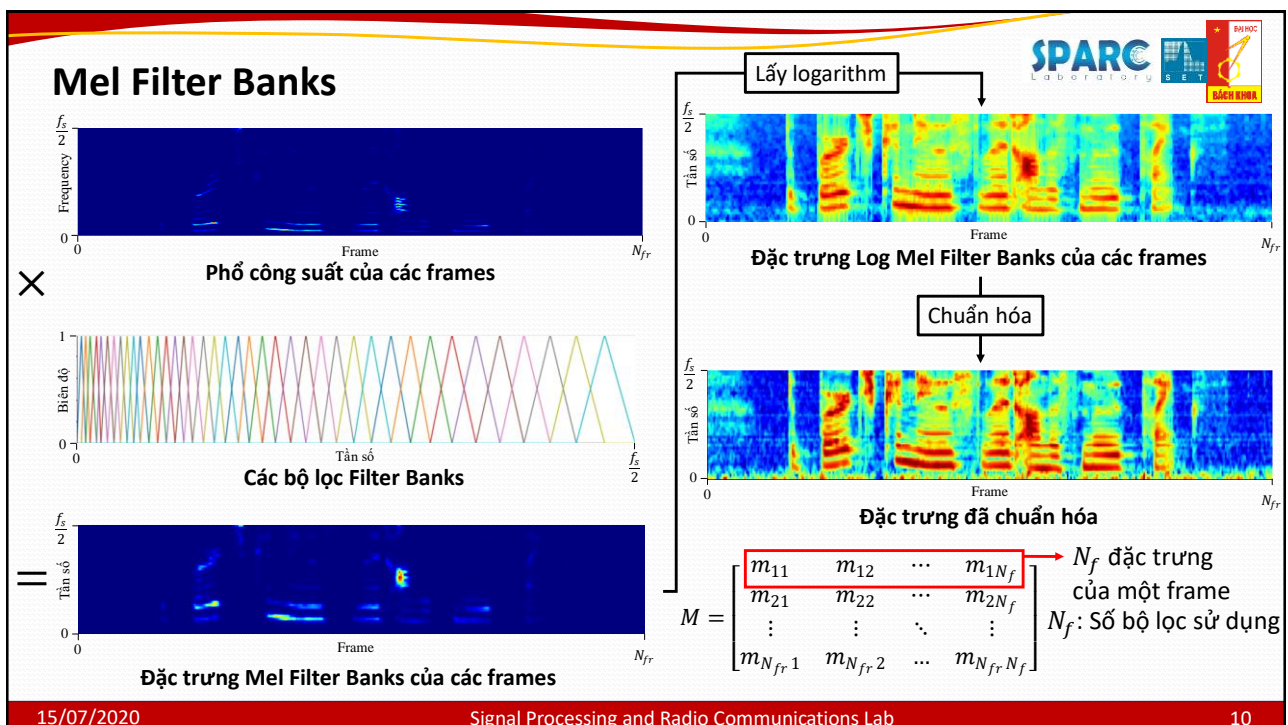
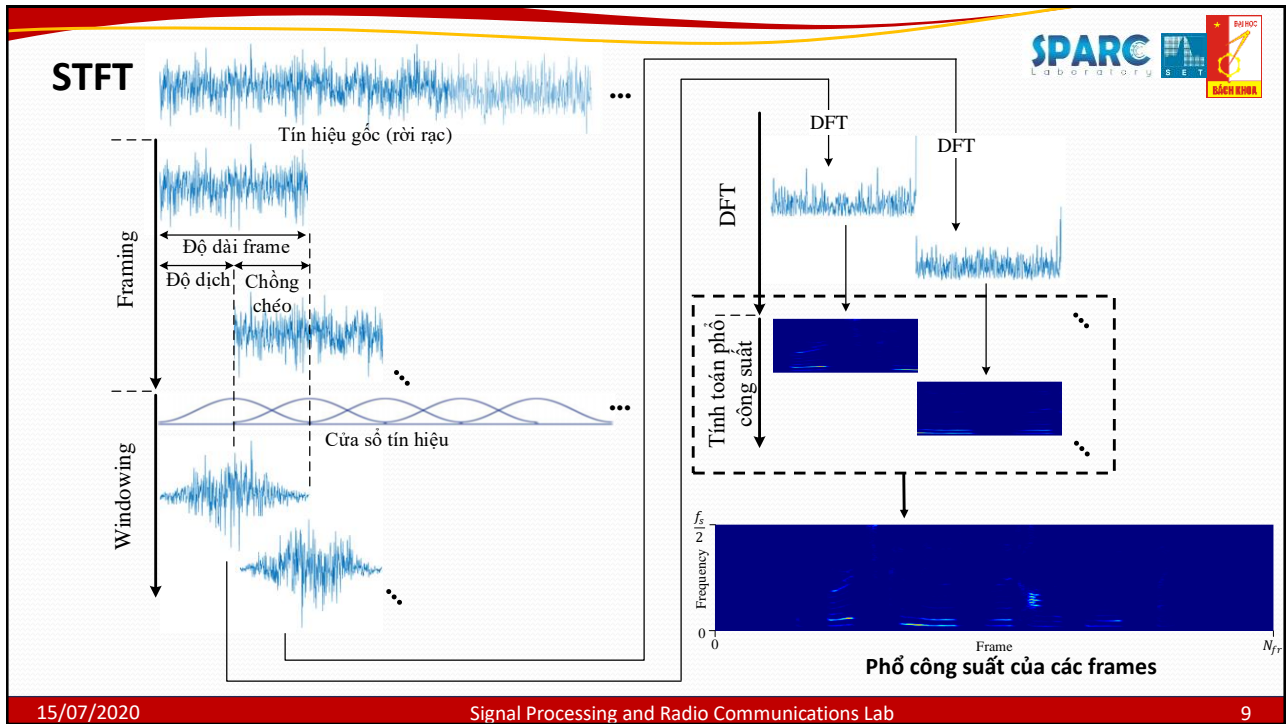


\* STFT: Short-time Fourier Transform

15/07/2020

Signal Processing and Radio Communications Lab

8



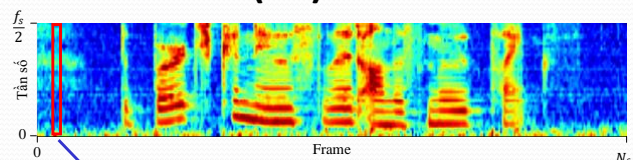


# Các mô hình mạng neural được sử dụng

SPARC  
Laboratory



## Autoencoder (AE)



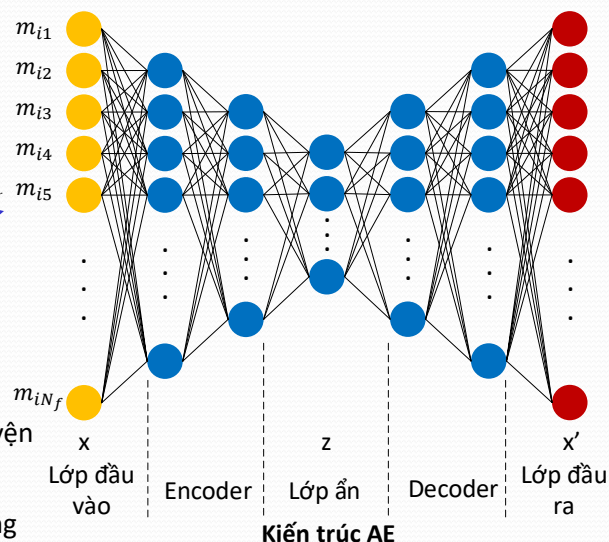
$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1N_f} \\ m_{21} & m_{22} & \dots & m_{2N_f} \\ \vdots & \vdots & \ddots & \vdots \\ m_{i1} & m_{i2} & \dots & m_{iN_f} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N_{fr}1} & m_{N_{fr}2} & \dots & m_{N_{fr}N_f} \end{bmatrix}$$

### Lỗi tái tạo:

$$\mathcal{L}(x, x') = \frac{1}{N} \|x - x'\|_2^2 \quad N: \text{Tổng số mẫu huấn luyện}$$

### Nguyên lý hoạt động:

- Huấn luyện AE chỉ sử dụng âm thanh bình thường
- $\mathcal{L}$  âm thanh bất thường >  $\mathcal{L}$  âm thanh bình thường



15/07/2020

Signal Processing and Radio Communications Lab

11

## Variational Autoencoder (VAE)

- Biểu diễn lớp ẩn bằng phân phối chuẩn  $\mathcal{N}(\mu, \sigma^2)$  thay vì bằng lớp neural cố định như AE
- Kỳ vọng ( $\mu$ ) và độ lệch chuẩn ( $\sigma$ ) được tính toán dưới dạng các neurons
- Lỗi tái tạo:

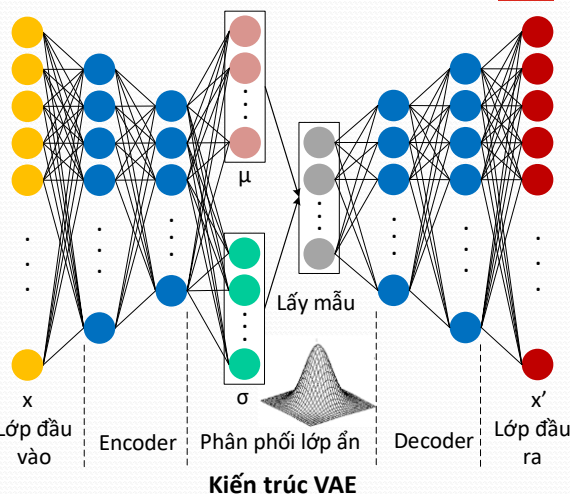
$$\mathcal{L}(x, x') = \frac{1}{N} (\|x - x'\|_2^2 + D_{KL})$$

N: Tổng số mẫu huấn luyện

$D_{KL}$ : Kulback-Leibler Divergence Loss

$D_{KL}$  thể hiện mất mát do sự xấp xỉ phân phối lớp ẩn với phân phối thực sự của nó

$$D_{KL} = -\frac{1}{2} (1 - \mu^2 - \sigma^2 + \log \sigma^2) [7]$$



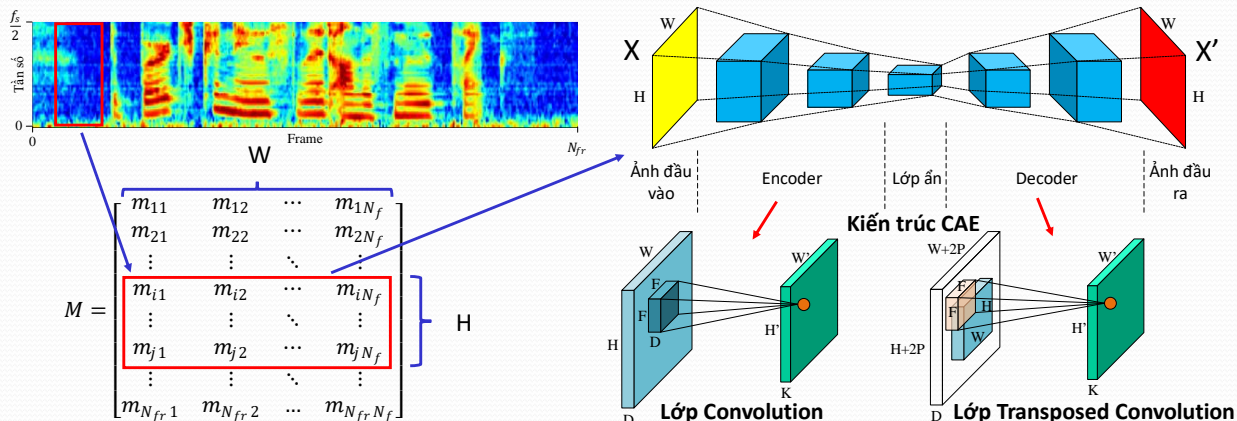
[7] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *International Conference on Learning Representations*, Banff, Canada, April 2014

15/07/2020

Signal Processing and Radio Communications Lab

12

## Convolutional Autoencoder (CAE)



- Ghép các đặc trưng của nhiều frame lại tạo thành ảnh hai chiều
- Các lớp Convolution và Transposed Convolution được sử dụng để xử lý dữ liệu dạng ảnh

### Lỗi tái tạo

$$\mathcal{L}(X, X') = \frac{1}{NHW} \|X - X'\|_2^2$$

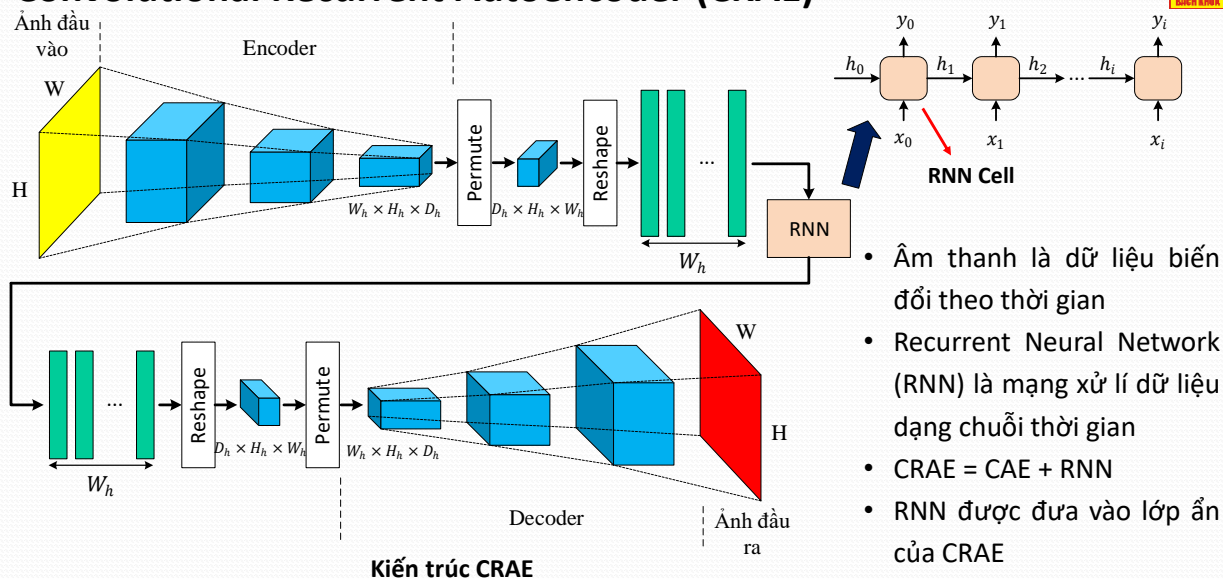
N: Tổng số mẫu huấn luyện

15/07/2020

Signal Processing and Radio Communications Lab

13

## Convolutional Recurrent Autoencoder (CRAE)



- Âm thanh là dữ liệu biến đổi theo thời gian
- Recurrent Neural Network (RNN) là mạng xử lý dữ liệu dạng chuỗi thời gian
- CRAE = CAE + RNN
- RNN được đưa vào lớp ẩn của CRAE

15/07/2020

Signal Processing and Radio Communications Lab

14

# Thí nghiệm và kết quả



## Tập dữ liệu

Tập dữ liệu MIMII [8]

	Máy bơm	Thanh ray trượt
Tập huấn luyện	816 segments bình thường	871 segments bình thường
Tập xác thực	90 segments bình thường	97 segments bình thường
Tập kiểm tra	143 segments bất thường, 100 segments bình thường	356 segments bất thường, 100 segments bình thường

## Các tham số đánh giá

### F-score (F1)

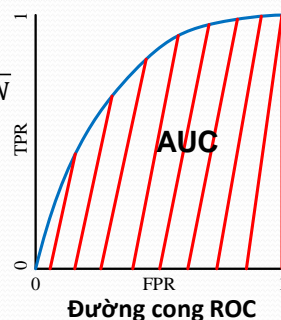
$$F1 = \frac{2TP}{2TP + FP + FN}$$

### ROC AUC

$$TPR = \frac{TP}{TP + FN}; FPR = \frac{FP}{FP + TN}$$

### Các tham số thống kê trung gian

	Nhãn	Dự đoán
True positive (TP)	Bất thường	Bất thường
False positive (FP)	Bình thường	Bất thường
True negative (TN)	Bình thường	Bình thường
False negative (FN)	Bất thường	Bình thường



Đường cong ROC

[8] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa and Y. Kawaguchi, "MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection," in Detection and Classification of Acoustic Scenes and Events 2019, New York, USA, October 2019

15/07/2020

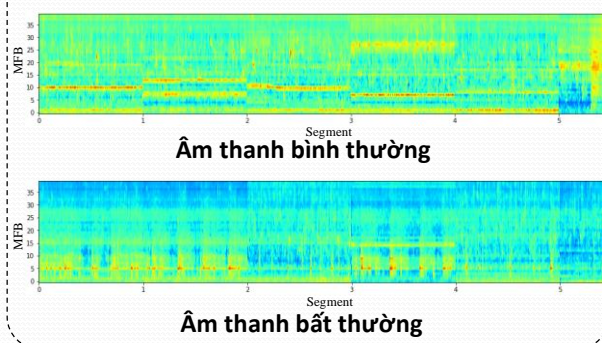
Signal Processing and Radio Communications Lab

15

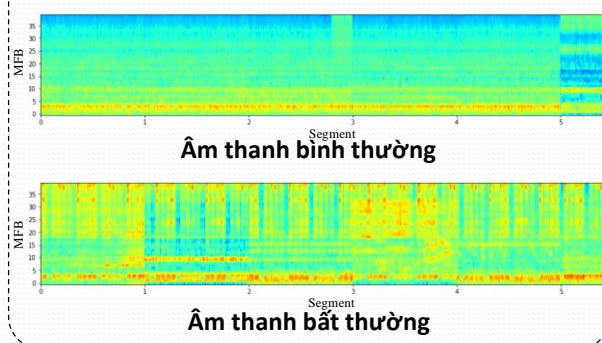
## Đặc trưng MFBs của tiếng máy



### MFBs cho tiếng máy bơm



### MFBs cho tiếng thanh ray trượt



- MFBs của âm thanh bình thường cho tiếng máy không có độ biến động cao theo thời gian
- MFBs của âm thanh bình thường là khác so với âm thanh bất thường

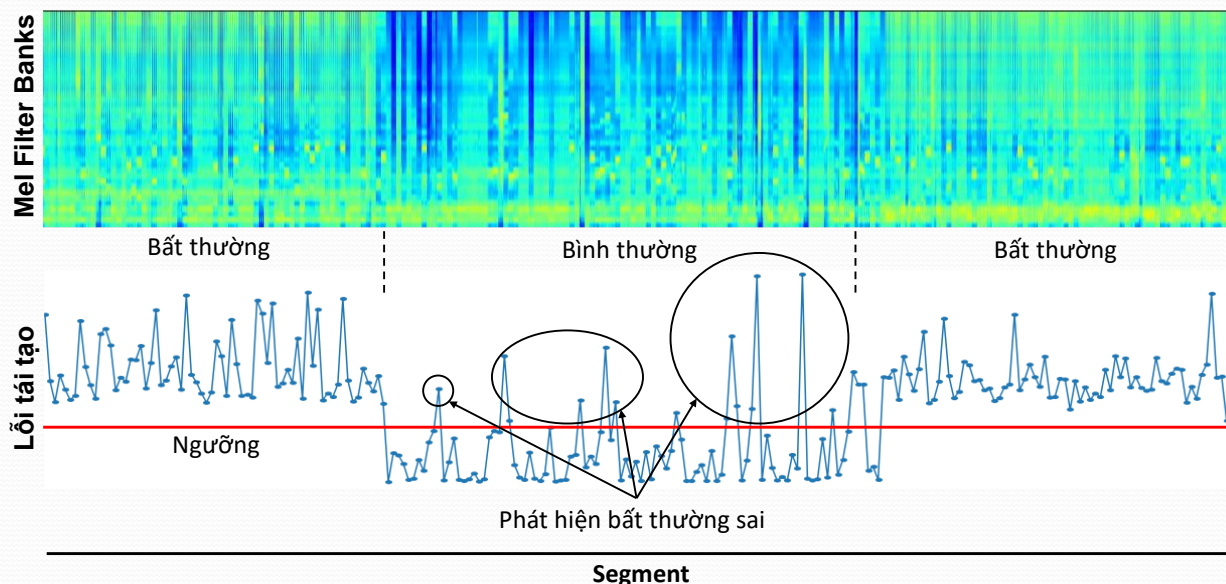
15/07/2020

Signal Processing and Radio Communications Lab

16



## Phát hiện bất thường bằng lỗi tái tạo



15/07/2020

Signal Processing and Radio Communications Lab

17

## Kết quả thí nghiệm

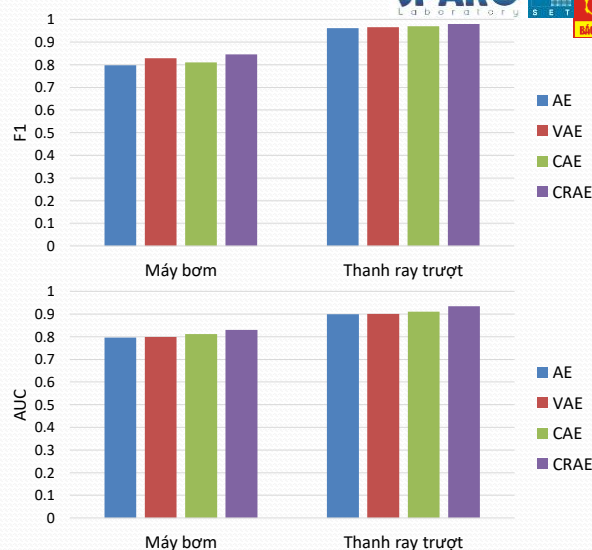
	Máy bơm		Thanh ray trượt	
	F1	AUC	F1	AUC
AE [4]	0.7971	0.796	0.9544	0.8994
VAE [4]	0.8231	0.7996	0.9654	0.9003
CAE [5]	0.8106	0.8115	0.9697	0.9113
CRAE [6]	0.8457	0.8308	0.9696	0.935

### • Tiếng máy bơm

- F1 và AUC tăng theo độ phức tạp của mô hình
- Hiệu quả của VAE và CAE khá tương đồng
- CRAE vượt trội so với các mô hình còn lại

### • Tiếng thanh ray trượt

- F1 tăng rất ít khi tăng độ phức tạp mô hình
- AUC của CRAE cao nhất



[4] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo and Y. Kawaguchi, "Anomalous Sound Detection Based on Interpolation Deep Neural Network," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, Barcelona, Spain, May 2020

[5] R. Müller, F. Ritz, S. Illium and C. Linhoff-Popien, "Acoustic Anomaly Detection for Machine Sounds based on Image Transfer Learning," in *arXiv:2006.03429 [eess.AS]*, June 2020

[6] S. R. Bukka, A. R. Magee and R. K. Jaiman, "Deep Convolutional Recurrent Autoencoders for Flow Field Prediction," *arXiv:2003.12147 [physics.flu-dyn]*, March 2020

15/07/2020

Signal Processing and Radio Communications Lab

18

# Kết luận

## Kết luận chung

- Autoencoder và các biến thể hoạt động hiệu quả trong phát hiện âm thanh bất thường
- CRAE có thể được áp dụng để giải quyết bài toán phát hiện âm thanh bất thường
- Độ hiệu quả của phát hiện bất thường được cải thiện theo độ phức tạp của mô hình
- Đối với tập dữ liệu MIMII, CRAE hoạt động hiệu quả nhất

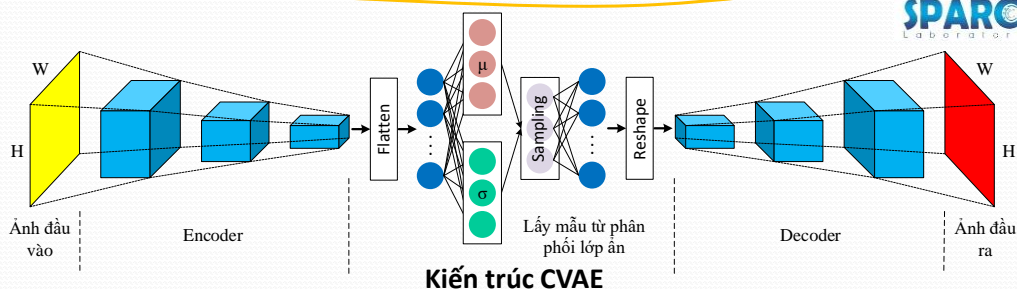
## Hướng phát triển

- Thử nghiệm thêm các mô hình khác, điển hình là Convolutional Variational Autoencoder (CVAE)
- Thí nghiệm trên tập những dữ liệu khác (âm thanh đường phố, âm thanh trong phòng)
- Ứng dụng phát hiện âm thanh bất thường thời gian thực bằng thiết bị Voice-IP

15/07/2020

Signal Processing and Radio Communications Lab

19



Thiết bị Voice-IP

15/07/2020

Signal Processing and Radio Communications Lab

20



**CẢM ƠN CÁC THẦY VÀ CÁC BẠN ĐÃ  
CHÚ Ý LẮNG NGHE!**