

## **Nâng cao khả năng phân loại tín hiệu thủy âm sử dụng nội suy khối nổi trục và phân phối xác suất trong không gian ẩn**

Bạch Nhật Hoàng, Nguyễn Trung Kiên\*, Vũ Lê Hà

Viện Khoa học và Công nghệ quân sự.

\*Email: kiennt67@gmail.com

Nhận bài: 11/4/2022; Hoàn thiện: 21/5/2022; Chấp nhận đăng: 15/6/2022; Xuất bản: 28/6/2022.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.80.2022.39-48>

### **TÓM TẮT**

*Trong lĩnh vực âm học đại dương, các phương pháp xử lý tín hiệu dưới nước truyền thống và hiện đại đều đang đạt được các kết quả tích cực. Đối với các bài toán sonar phục vụ nhiệm vụ quốc phòng an ninh, nhu cầu phân loại kịp thời, chính xác các dạng tàu chân vịt là một yêu cầu quan trọng hàng đầu. Nghiên cứu này trình bày một mô hình xử lý tín hiệu dưới nước nhằm mục đích phát hiện và phân loại tàu chân vịt thông qua cách cải tiến kỹ thuật LOFAR (Low Frequency Analysis and Recording) bằng nội suy khối nổi trục (Cubic Splines Interpolation - CSI), kết hợp với mô hình xác suất phân phối trong miền không gian ẩn. Kết quả của mô hình được thử nghiệm trên các tập dữ liệu thực, đã cho thấy độ chính xác phân loại tốt hơn 10%, đạt hiệu suất khoảng 90% so với các mô hình trước đó. Giải pháp này cũng đã chứng minh được mô hình kết hợp truyền thống và hiện đại có khả năng phân loại tốt với các tín hiệu thực tế ngay cả khi lượng dữ liệu thiếu hụt và tỷ lệ tín trên tạp thấp.*

**Từ khoá:** Tín hiệu thủy âm; Nội suy khối nổi trục; Phân phối xác suất; Không gian ẩn.

### **1. MỞ ĐẦU**

Hệ thống sonar (Sound navigation and ranging) sử dụng các lý thuyết về sóng âm trong môi trường nước để truyền tin, phát hiện, định vị và theo dõi mục tiêu [1, 2]. Để phục vụ bài toán phân loại đối tượng nguồn âm dưới nước, hai loại phân tích điển hình được thực hiện để trích xuất các đặc trưng của tín hiệu là: phương pháp tách sóng đường bao DEMON (Detection Envelope Modulation On Noise) [2], và Phương pháp phân tích tần số LOFAR (Low Frequency Analysis and Recording) [3]. LOFAR là một kỹ thuật phân tích băng rộng, cho ta các thông tin ước lượng về độ rung của động cơ đẩy, cũng như các tần số dao động sinh ra từ các máy móc cơ khí trên tàu, từ đó tìm ra đặc trưng của đối tượng [4]. Gần đây cùng với sự phát triển của lĩnh vực trí tuệ nhân tạo, vào năm 2019, các nhà khoa học đã chứng minh được việc sử dụng mô hình mạng Nơ-ron (Neural Network-NN) có thể tiếp tục cải thiện kết quả phân loại của các mô hình dưới nước [5]. Với đặc trưng của môi trường thủy âm, các tín hiệu từ các nguồn sinh học (như các động vật biển...) hoặc phi sinh học (tàu thuyền, gió, sét, rung động địa chấn,...) có thể thay đổi tùy thuộc vào điều kiện môi trường hoạt động, thời gian,...; điều này ảnh hưởng đến sự ổn định việc xử lý tín hiệu theo các nguyên lý sonar. Do đó, những thay đổi trong số liệu thống kê của các tín hiệu sonar phải luôn được theo dõi trong thời gian thực. Các phương pháp phân loại sử dụng học máy (Machine Learning-ML) nói chung và học sâu (Neural Network-NN) nói riêng có khả năng thực hiện tự động các nhiệm vụ đó với chất lượng kết quả tương đối cao. Do đó, xu hướng kết hợp giữa các kỹ thuật xử lý tín hiệu truyền thống và mô hình trí tuệ nhân tạo là một cách tiếp cận hứa hẹn nhiều nghiên cứu và đột phá trong tương lai. Nhìn chung, tín hiệu sonar được đưa vào các thuật toán tiền xử lý để trích xuất đặc trưng, giảm kích thước và tăng cường khả năng biểu diễn của dữ liệu. Dữ liệu đã xử lý này sẽ là đầu vào của một mạng học sâu (Deep Learning-DL) [5, 6]. Các nghiên cứu sử dụng mô hình mạng Nơ-ron xác suất (Probabilistic Neural Network-PNN) [7], Lượng tử hóa vector (Learning Vector Quantization-LVQ) [8], hoặc các dạng mạng học sâu cơ bản (Multilayer Perceptron-MLP) [9], đến phức tạp (Deep Neural Network-DNN) [10] cũng bước đầu cho các kết quả tích cực. Trong những năm gần đây, việc sử

dụng Mạng Nơ-ron tích chập (Convolutional Neural Network-CNN) cũng đã thu được những kết quả đáng kể trong các mô hình phân loại tín hiệu dưới nước [6, 10]. Tuy nhiên, vẫn chưa có nhiều các kết quả công bố về giai đoạn tiền xử lý dữ liệu thủy âm để nâng cao chất lượng dữ liệu trước khi sử dụng làm đầu vào cho mô hình trí tuệ nhân tạo. Dựa trên các phân tích trên, các phần tiếp theo của bài báo được trình bày như sau: Phần 2 trình bày cải tiến LOFAR bằng nội suy khối nổi trực để xây dựng bộ dữ liệu đầu vào cho mô hình đề xuất; Phần 3 trình bày giải pháp cải tiến mô hình phân phối xác suất, sau đó kết hợp hai giải pháp đề xuất và các kết quả phân loại trên tập dữ liệu thực; Và cuối cùng là kết luận chung.

## 2. TIỀN XỬ LÝ DỮ LIỆU BẰNG NỘI SUY KHỐI NỔI TRỰC

### 2.1. Kỹ thuật LOFAR

#### 2.1.1. LOFAR truyền thống

Giả sử một mảng có  $N$  phần tử, chuỗi tín hiệu đầu vào là  $x_i(k)$  với  $i=1,2,\dots,M$ . Tín hiệu được định hướng theo hướng  $m$  được biểu thị bằng  $x_i(k,m)$  với  $m=1,2,\dots,M$ . Trong đó,  $M$  là số lượng chùm tia và  $\theta_0$  là góc của các chùm tia liên kế được tính bằng công thức  $\theta_0 = \frac{3600}{M}$ .

Tín hiệu chùm tia thứ  $m$  là:

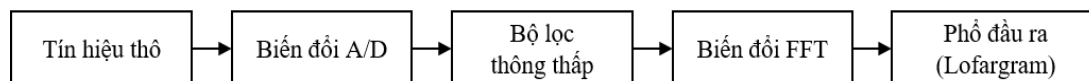
$$y_m(k) = \sum_{i=1}^N x_i(k,m) \quad (1)$$

Tín hiệu tại đầu thu là tín hiệu sóng cơ, do đó có dạng:

$$y(t) = A \cos(2\pi f_0 t + \varphi) + n(t) \quad (2)$$

Trong đó,  $n(t)$  là nhiễu,  $\varphi$  là pha ngẫu nhiên,  $A$  là biên độ, tỉ số tín trên tạp SNR là  $\frac{A^2}{2\sigma_n^2}$ .

Do đó, đối tượng chính cần phân tích của LOFAR sẽ là tín hiệu đầu vào  $y_m(k)$ . Giá trị đầu ra của chùm tia là  $D_m = E[y_m^2(k)]$ . Nhìn chung,  $D_m$  thay đổi chậm và chứa thông tin về góc tới của tín hiệu cần tìm và dữ liệu  $y_m(k)$  cho ta các thông tin về tín hiệu mục tiêu, nhiễu, sự giao thoa.



Hình 1. Sơ đồ khối của kỹ thuật LOFAR truyền thống [3].

Vì LOFAR là phép phân tích trên miền tần số, do đó, kỹ thuật biến đổi Fourier nhanh (Fast Fourier transform-FFT) sẽ đóng vai trò rất quan trọng. Kết quả cuối cùng của phép phân tích LOFAR giúp ta phát hiện ra các tần số sóng đơn dưới dạng sóng sin. Do đó, tham số quan trọng nhất của tín hiệu này là biên độ và tần số.

#### 2.1.2. Hạn chế của LOFAR truyền thống và cách khắc phục

Trong thực tế, các tín hiệu thủy âm là các tín hiệu biến đổi liên tục và phức tạp, các tần số đặc trưng xuất hiện trong môi trường không phải luôn là các tín hiệu tuần hoàn. Do đó, sử dụng biến đổi Fourier nhanh sẽ gặp phải hạn chế về độ phân giải trong miền thời gian. FFT cho chúng ta biết chính xác tần số nào xuất hiện trong một tín hiệu, nhưng không biết thời điểm nào mà các tần số này đã xuất hiện trong tập dữ liệu. Do đó, ta không thể biết tần số phát hiện được xảy ra trong thời gian nào của dữ liệu và cũng không thể phân biệt cụ thể vậy tín hiệu đó liệu có thuộc về nguồn âm cần tìm hay không. Để khắc phục vấn đề này, ta có thể thay thế biến đổi FFT bằng biến đổi Fourier thời gian ngắn (Short time Fourier transform-STFT). Phép biến đổi STFT thêm chiều thời gian vào các tham số của hàm cơ sở bằng cách nhân hàm mũ phức vô hạn với một cửa sổ.

$$b(\omega, t_0)(t) := \omega(t - t_0) \exp(i \omega t) \tag{3}$$

Kết quả của phép biến đổi này cũng có thể được coi là một bộ lọc thông dải có biến đổi fourier với cửa sổ  $\omega(t)$  là đáp ứng tần số được dịch theo tần số trung tâm  $\omega$ . Do đó, tất cả các bộ lọc có cùng băng thông. Ta có công thức tổng quát cho STFT như sau:

$$S\{s(t)\}(\omega, t_0) = \int_{-\infty}^{\infty} \omega(\tau) * \exp(-i \omega \tau) s(\tau) d\tau \tag{4}$$

Từ đó, ta có thể xác định thời điểm nào thì tần số nào xuất hiện, từ đó cải thiện được độ phân giải theo chiều thời gian của tín hiệu. Mô hình đề xuất của bài báo, đầu tiên sử dụng cách tiếp cận giống như trong [11] đã đề xuất. Bằng cách hạ tần xuống  $\delta$  lần để tránh xung răng cưa, loại bỏ các tần số không mong muốn, kết quả lúc đó sẽ thu được tần số  $f_s / \delta$ . Sau đó, nhân tần số này với cửa sổ Hanning để tạo ra các phân đoạn tín hiệu; trước khi các phân đoạn tín hiệu được biến đổi sang miền tần số bằng STFT. Tại đây, ta sử dụng cách xếp chồng các phân đoạn dữ liệu với tỷ lệ 60% để tăng tính liên tục của các mẫu tín hiệu, làm tăng độ chính xác của LOFAR.

Tóm lại, bằng việc thay thế biến đổi FFT thành STFT, cách tiếp cận mới sẽ biến đầu ra của phân tích LOFAR cũ thành LOFARgram với trục x là tần số và trục y là thời gian. Tuy nhiên, do điều kiện các bộ dữ liệu về tín hiệu thủy âm là rất khan hiếm và khó tiếp cận, các tín hiệu sinh ra trong quá trình di chuyển của tàu cũng thay đổi phức tạp, có sự chồng lấn về các dải tần, do đó, ta cần chuẩn hóa dữ liệu sau khi biến đổi STFT để tăng thông tin có ích trên các ảnh phổ Lofargram.

**2.2. Giải pháp sử dụng nội suy khối nối trục**

Trong thực tế, khi xây dựng một bộ dữ liệu, quá trình thu và lưu trữ tín hiệu thủy âm của các đối tượng nguồn âm đòi hỏi rất nhiều tham số về điều kiện môi trường, thời điểm triển khai hệ thống, độ ổn định của các cấu trúc phần cứng. Do đó, tính ổn định của một bộ dữ liệu có thể không chắc chắn. Bên cạnh đó, khi xử lý các bản thu dữ liệu thô, bằng các phương pháp xử lý tín hiệu số sử dụng các tham số như độ dài cửa sổ trượt, các hệ số bộ lọc v.v. sẽ dẫn đến các kết quả sai lệch khi phân loại do sự khác nhau về độ phân giải của từng phép toán với nhau. Do đó, ta có thể áp dụng thuật toán nội suy để nâng cao chất lượng của các ảnh phổ sau khi biến đổi từ các dữ liệu thủy âm thô, trước khi đưa vào mô hình phân loại.

Hầu hết các tín hiệu thủy âm đều có các đặc tính chung của các chuỗi thời gian là khi số lượng và cường độ tần số tăng lên thì tỷ lệ SNR sẽ giảm đi, do đó, lượng thông tin hữu ích dùng để phân loại đối tượng mà ta có thể bóc tách được từ dữ liệu thô ban đầu giảm xuống. Vấn đề càng khó khăn hơn đối với các bộ dữ liệu thủy âm không công khai cách đo đạc, không mô tả quá trình thu tín hiệu thô, do đó, để giảm thiểu, các nghiên cứu về thủy âm có thể sử dụng một số bộ lọc thông kê là bộ lọc trung bình động [12, 13]. Tuy nhiên, cách tiếp cận này sẽ làm thay đổi cấu trúc tương quan của dữ liệu thậm chí có thể làm mất một số điểm dữ liệu ở đầu hoặc cuối chuỗi dữ liệu của tín hiệu thô. Để tránh sự bất ổn khi khai triển bậc đa thức của nội suy đa thức bậc cao và các hạn chế về mô hình thống kê, ta có thể sử dụng nội suy khối nối trục [14, 15]. Dữ liệu đã tiền xử lý được biểu diễn dưới dạng ảnh phổ LOFARgram được đưa qua một bộ nội suy khối nối trục để ước tính các giá trị pixel trung gian giữa các giá trị đã biết để nâng cao chất lượng thông tin chứa đặc trưng của ảnh phổ. Từ công thức (4) tổng quát cho khai triển STFT, ta có phương trình STFT của một tín hiệu  $x(n)$  với cửa sổ trượt  $w(n)$  được cho bởi công thức:

$$STFT\{x[n]\} \equiv X(p, w) = \int_{p=-\infty}^{+\infty} x[p]w[p-n]e^{-j\omega p} \tag{5}$$

Các biến đổi nội suy tuyến tính nối trục (Linear Interpolation splines) và nội suy khối nối trục

(Cubic Interpolation splines) cho các điểm nằm giữa hai điểm biên  $(I_K, V_K)$  và lần lượt được cho bởi công thức (6) và (7):

$$I_K(V) = I_K + \frac{I_{K+1} - I_K}{V_{K+1} - V_K} (V - V_K) \quad (6)$$

$$I_K(V) = I_K + a_K(V - V_K) + b_K(V - V_K)^2 + c_K(V - V_K)^3 \quad (7)$$

Khi đó, áp dụng công thức (5) vào công thức (6) và (7), ta có giá trị các điểm lân cận giữa 2 điểm  $p$  và  $p_c$  được cho bởi công thức (8) và (9):

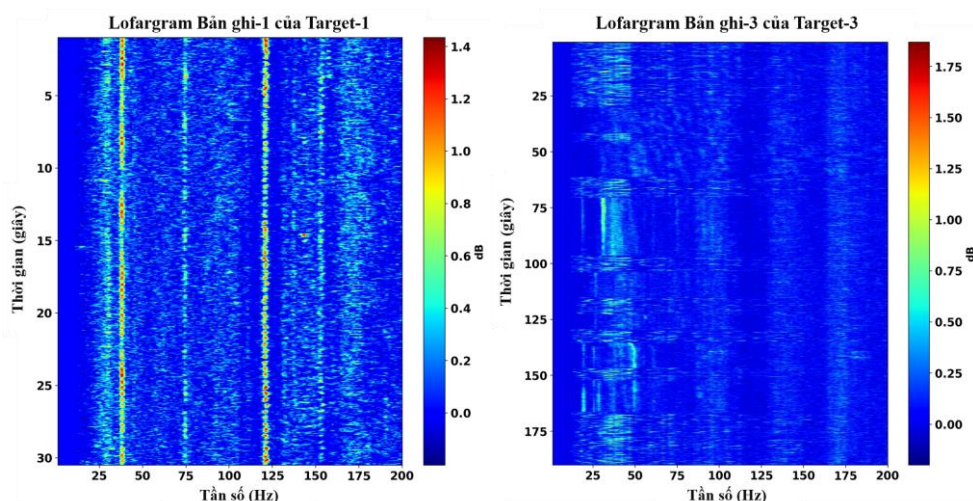
$$w = w_i + \frac{w_{i+1} - w_i}{n_{i+1} - n_i} (p - p_i) \quad (8)$$

$$w = w_i + a_K(p - p_i) + b_K(p - p_i)^2 + c_K(p - p_i)^3 \quad (9)$$

Với  $i = 1, 2, \dots, c-1$ ,  $c$  là số điểm theo hệ số cho trước.  $a_k, b_k, c_k$  là các hệ số nội suy nối trục.

Từ công thức (8) và (9), ta có thể thấy giải pháp nội suy khối nối trục đã khắc phục được các hạn chế về mặt toán học của các giải pháp nội suy đơn điểm (simple interpolation traditional), nội suy tuyến tính (linear interpolation) đó là do, các mô hình bậc 3 của nội suy khối nối trục đã tạo ra các hàm bậc ba có đạo hàm bậc nhất ở các điểm nối dữ liệu bằng nhau, tạo ra sự liên tục và tính kế thừa của các điểm dữ liệu được nội suy giữa các điểm dữ liệu đã có.

Trong khi đó, với giải pháp nội suy đơn điểm, các giá trị được nội suy ở các thời điểm ngay trước khi thống kê sẽ trở nên gián đoạn và không liên tục; với nội suy tuyến tính, tuy đã khắc phục được sự gián đoạn nhưng tại các điểm nội suy, đạo hàm bậc nhất lại không liên tục. Rõ ràng, với các tín hiệu có độ thay đổi phức tạp về mặt tần số như tín hiệu thủy âm, việc sử dụng các phương pháp thống kê bậc cao như nội suy nối trục là một hướng nghiên cứu khả thi có thể sử dụng tương đương một bộ lọc tín hiệu số để nâng cao khả năng trích xuất đặc trưng của giai đoạn tiền xử lý.



Hình 2. LOFARgram của Target-1 và Target-3.

Thuật toán nội suy khối nối trục cho kỹ thuật LOFAR sẽ được triển khai như sau:

**Đầu vào:** Tín hiệu  $x$ , hàm  $w$ , và bộ hệ số  $\theta = [\theta_1, \theta_2, \dots, \theta_n]$

**Đầu ra:** Một tensor  $\mathbb{Z}$  với  $l$  chiều

1 Khởi tạo giá trị nhỏ nhất của  $w_0$  và  $n_0$

- 2 Nếu  $i = 1$  tới  $k$  thì thực hiện
- 3 Tính toán giá trị Spectrogram  $S_i = STFT(x, w, \theta_i)$  (Công thức 4)
- 4 Giới hạn biên giá trị của  $w_0$  và  $n_0$  luôn là nhỏ nhất
- 5 Nếu  $\Delta w_i < w_0$  thì
- 6  $w_0 = \Delta w_i$
- 7 Kết thúc
- 8 Nếu  $\Delta n_i < n_0$  thì
- 9  $n_0 = \Delta n_i$
- 10 Kết thúc
- 11 Kết thúc
- 12 Nếu  $i = 1$  tới  $k$  thì thực hiện
- 13 Nội suy từng Spectrogram  $CS_i = INTERPOLATE(S_i, w_0, n_0)$  (Công thức 8)
- 14 Kết thúc
- 15 Xếp chồng các ảnh phổ đã nội suy  $Z = [CS_1, CS_2, \dots, CS_k]$

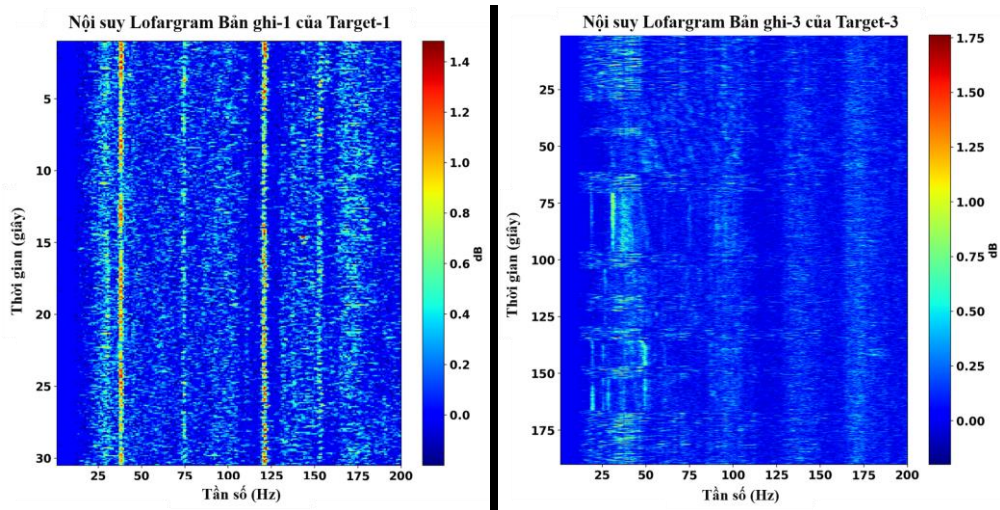
Các dữ liệu chuẩn dùng trong bài báo được lấy từ tập dữ liệu thu thực tế trong quá trình hoạt động trong môi trường biển Việt Nam của bốn lớp tàu động cơ Diesel, dữ liệu của bốn lớp tàu này cũng được thu trong các điều kiện ổn định khác nhau về môi trường, thời gian, tốc độ di chuyển. Từ dữ liệu thực tế sau khi sử dụng kỹ thuật LOFAR, các dải tần số đặc trưng tập trung chủ yếu trong khoảng 25 Hz và 75 Hz.

Do đó, khai triển phương trình (9) để đưa ra các hệ số nội suy khối nối trực sử dụng trong nghiên cứu này, được cho bởi công thức (10) và (11):

$$w = x^3 - 2x^2 + 1, \quad \text{nếu } 10 \leq x \leq 25 \quad (10)$$

$$w = 3x^3 + 5x^2 - 8, \quad \text{nếu } 25 \leq x \leq 75 \quad (11)$$

Kết quả của mô hình LOFAR và LOFAR nội suy được biểu diễn trong hình 2 và hình 3. Các ảnh phổ sau khi đi qua phép nội suy sẽ được đưa vào một bộ phân loại sử dụng phương pháp phân phối xác suất để đưa ra kết quả và đánh giá độ chính xác của mô hình đề xuất tại phần 3.



**Hình 3.** LOFARgram sau khi nội suy của Target-1 và Target-3.

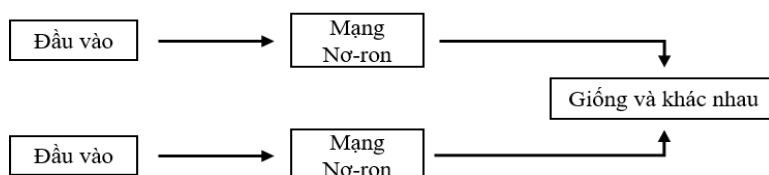
### 3. PHÂN LOẠI TÍN HIỆU BẰNG PHÂN PHỐI XÁC SUẤT TRONG KHÔNG GIAN ẨN

Các mô hình NN truyền thống luôn đòi hỏi lượng lớn dữ liệu để huấn luyện mô hình, mà

trong thực tế, có nhiều vấn đề mà việc thu thập đủ dữ liệu là một trở ngại lớn, ví dụ như xây dựng bộ dữ liệu thủy âm, do đó, Siamese Network (SN) là một giải pháp để khắc phục tình trạng thiếu hụt dữ liệu đó. Khả năng học các đặc trưng từ lượng dữ liệu nhỏ đã làm cho SN trở nên phổ biến hơn trong những năm gần đây.

### 3.1. Siamese Network (SN) và Variational Auto-Encoder (VAE)

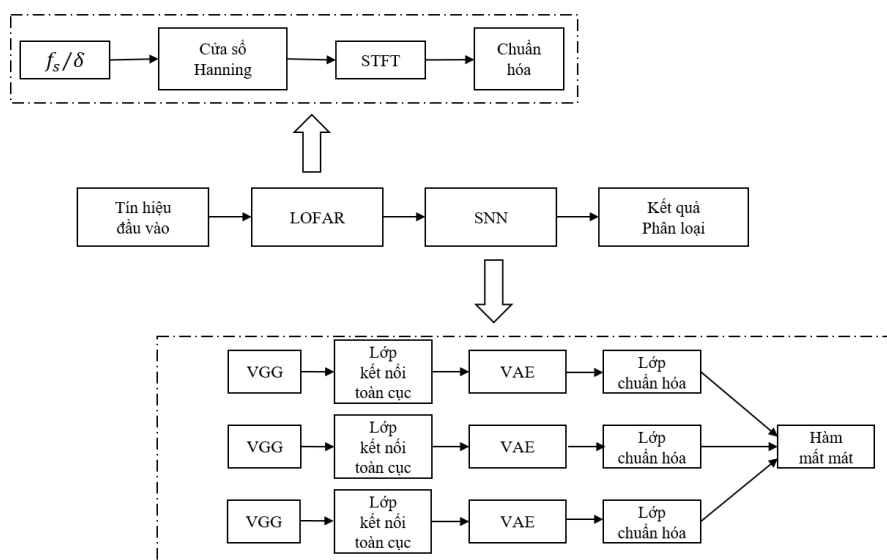
SN [16] là một kiến trúc mạng Nơ-ron chứa hai hoặc nhiều mạng con có cùng cấu hình, cùng thông số và các hàm trọng số. Việc cập nhật các thông số được thực hiện đồng thời trên tất cả các mạng con của SN. SN được sử dụng để tìm sự giống nhau của các dữ liệu đầu vào bằng cách so sánh các vectơ đặc trưng của chúng. Với các NN truyền thống, khi xuất hiện thêm các đối tượng mới, ta thường phải huấn luyện lại trên toàn bộ tập dữ liệu (bao gồm cả dữ liệu mới và cũ). Ngoài ra, các mạng Nơ-ron sâu cần một khối lượng lớn dữ liệu để có thể huấn luyện chúng. SN, theo một cách khác, học cách tìm ra sự giống nhau giữa các dữ liệu đầu vào. Vì vậy, nó cho phép chúng ta phân loại các lớp dữ liệu mới mà không cần huấn luyện lại mạng Nơ-ron.



Hình 4. Mô hình ý tưởng của mạng Siamese.

VAE là một dạng autoencoder dựa trên xác suất, trong đó, các đầu ra xuất hiện dưới dạng các phân phối xác suất. Các mô hình AutoEncoder thông thường có xu hướng tập trung làm giảm tối đa có thể hàm mất mát (hàm mất mát), do đó, với các tín hiệu phức tạp sẽ dẫn đến một không gian ẩn có sự bất thường. VAE có thể khắc phục được điều đó bằng cách thay vì encode đầu vào như một điểm duy nhất, VAE sẽ encode đầu vào dưới dạng phân phối trong không gian ẩn và chuẩn hóa ma trận hiệp phương sai và trung bình của phân phối.

### 3.2. Mô hình đề xuất kết hợp SNN-VAE



Hình 5. Mô hình phân loại tín hiệu thủy âm đề xuất.

Mô hình đề xuất được mô tả trong hình 5, với việc sử dụng cấu trúc mạng triple mất mát, ta có giá trị hàm mất mát của mô hình SNN-VAE được tính bởi một hàm là tổng giá trị suy hao của mạng SN và mô hình phân phối xác suất trong không gian ẩn.

$$L_{SNN-VAE} = L(A, P, N) + L_{reconstruct} + \beta KL(z, N(0, I_d)), \quad (12)$$

Với  $L(A, P, N)$  là hàm suy hao của mạng SN,  $L_{reconstruct}$  là hàm mất mát khi tái cấu trúc lượng thông tin khi decode, và  $\beta KL(z, N(0, I_d))$  là hàm mất mát khi tái lập tính liên tục của dữ liệu.

Từ mô hình mạng Nơ-ron tích chập VGG sử dụng trong hình 5, mô hình sẽ đưa ra 2 véc tơ đã được encoding là  $x_1$  và  $x_2$  chứa các thông tin biểu diễn của lần lượt ảnh phổ thứ 1 và ảnh phổ thứ 2;  $x_1$  và  $x_2$  có cùng số chiều. Hàm  $f(x)$  có tác dụng tương tự như một phép biến đổi qua lớp kết nối toàn cục trong mạng NN để tạo tính phi tuyến và giảm chiều dữ liệu về các kích thước nhỏ.

Khi  $x_1$  và  $x_2$  là cùng một đối tượng hoặc không cùng là một đối tượng thì giá trị của  $\|f(x_1) - f(x_2)\|^2$  sẽ tương ứng là một giá trị nhỏ hoặc lớn. Mô hình sử dụng hàm mất mát là hàm triple loss có khả năng học sự giống nhau giữa hai ảnh phổ cùng nhóm và phân biệt các ảnh phổ không cùng nhóm. Mục tiêu của hàm mất mát là tối thiểu hóa khoảng cách giữa hai ảnh khi chúng khác nhau (negative) và tối đa hóa khoảng cách khi chúng là giống nhau (positive). Như vậy, chúng ta cần lựa chọn các bộ 3 ảnh được gọi là ảnh Anchor, ảnh Positive và ảnh Negative, sao cho:

- Ảnh Anchor và Positive khác nhau nhất: cần lựa chọn để khoảng cách  $d(A, P)$  lớn. Ta lựa chọn những ảnh phổ của cùng một đối tượng nhưng được ghi tại các thời điểm khác nhau để tạo thành cặp, cách sắp xếp này sẽ làm mô hình học tốt hơn.
- Ảnh Anchor và Negative giống nhau nhất: cần lựa chọn để khoảng  $d(A, N)$  nhỏ. Điều này tương tự như việc thuật toán phân biệt được ảnh phổ của 2 đối tượng tàu hải quân dù có cùng dải tần hoạt động nhưng khác nhau về cường độ và phân phối.

Mục đích cuối cùng là đạt được giá trị  $d(A, P) < d(A, N)$ , tức là ta có:

$$\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 \leq 0, \quad (13)$$

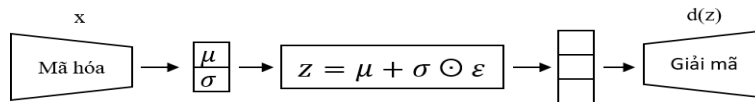
Để tránh cho mạng SNN-VAE sẽ mã hóa tất cả các đầu ra đều bằng nhau dẫn đến hàm  $f$  đầu ra luôn là 0, ta thêm vào 1 giá trị biên  $\alpha$  để hàm  $f$  có đầu ra nhỏ hơn 0 một giá trị đủ nhỏ.

$$\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha \leq 0, \quad (14)$$

Ta có hàm mất mát cho đầu ra của mạng SNN sẽ là:

$$L(A, P, N) = \sum_{i=0}^n \max(\|f(A_i) - f(P_i)\|^2 - \|f(A_i) - f(N_i)\|^2 + \alpha, 0), \quad (15)$$

với  $n$  là số bộ 3 ảnh được đưa vào để huấn luyện mô hình.



**Hình 6.** Mô hình phân phối xác suất trong không gian ẩn (VAE).

Hình 6 miêu tả một quá trình khi đầu vào  $x$  được xử lý trong không gian ẩn và cho kết quả là đầu ra  $d(z)$ , khi đó hàm mất mát của không gian ẩn trong VAE là:

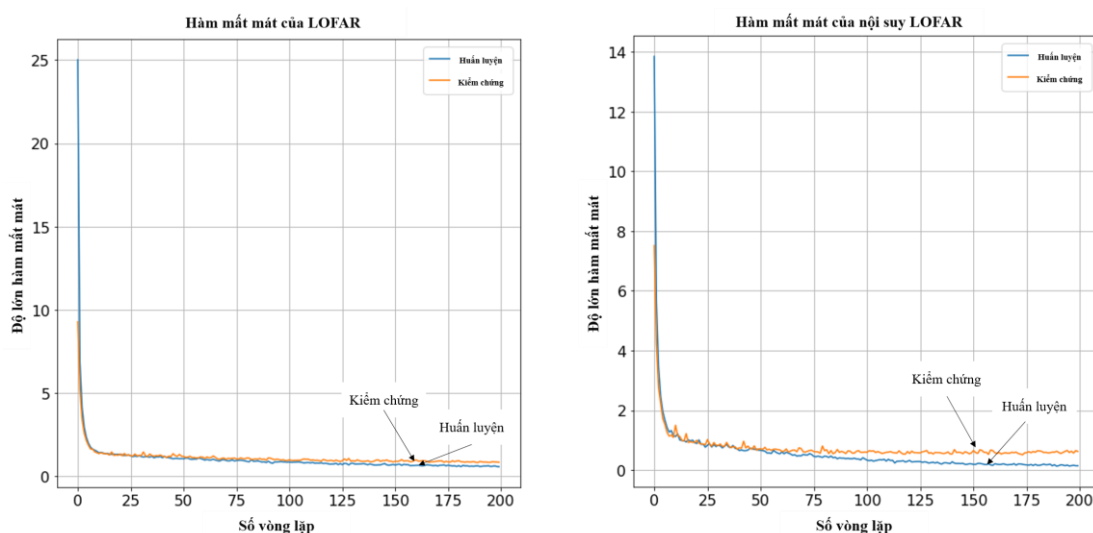
$$L_{VAE} = \|x - d(z)\|^2 + \beta KL(z, N(0, I_d)), \quad (16)$$

$\beta KL(z, N(0, I_d))$  là phép đo khoảng cách giữa 2 phân phối xác suất, sự khác nhau giữa hai phân phối xác suất quyết định sự khác nhau, và nếu phép đo có giá trị bằng 0 thì hai phân phối

được coi là giống nhau. Hàm mất mát sẽ cố gắng giảm thiểu sự phân kỳ giữa phân phối ban đầu và phân phối được tham số hóa, từ đó ngăn cản mạng học các phân phối hẹp và cố gắng đưa phân phối gần hơn với phân phối chuẩn đơn vị với  $E(z)$  và  $V(z)$  lần lượt là trung bình và hiệp phương sai của phân phối chuẩn.

$$\beta KL(z, N(0, I_d)) = \frac{1}{2} \sum_{i=1}^d (V(z_i) - \log V(z_i) - 1 + E(z_i)^2), \quad (17)$$

Khi đó, ta so sánh giá trị hàm mất mát của mô hình theo công thức (9) khi sử dụng hai bộ dữ liệu đầu vào lần lượt là spectrogram từ kỹ thuật LOFAR và spectrogram từ LOFAR cải tiến, để đánh giá hiệu quả khi phân loại bốn lớp tàu với nhau. Từ hình 7, ta có thể thấy hàm mất mát của mô hình phân loại với bộ dữ liệu đã nội suy có sự cải thiện so với bộ dữ liệu ban đầu, tránh được hiện tượng quá khớp trong khi huấn luyện, cải thiện được độ chính xác của mô hình. Tại đây, ta có thể thấy giá trị hàm mất mát của mô hình đã hội tụ hơn, giá trị của hàm mất mát tại các vòng lặp 100, 150 và 200 cho kết quả giá trị loss của bộ dữ liệu LOFARgram sau khi nội suy đã giảm đáng kể so với không nội suy, điều đó đã đảm bảo tính hội tụ của mô hình. Ta sử dụng cấu hình của mạng VGG truyền thống theo [17] với các cửa sổ chập (3x3), lớp gộp cực đại (2x2) có bước sai là 2, các khối chập của mạng VGG lúc này sẽ gồm một chuỗi các tầng có cấu trúc tương tự nhau như sau: 1 lớp tích chập có chèn để giữ độ phân giải ổn định, 1 lớp phi tuyến sử dụng hàm kích hoạt RELU, và 1 lớp gộp cực đại để giảm chiều không gian.

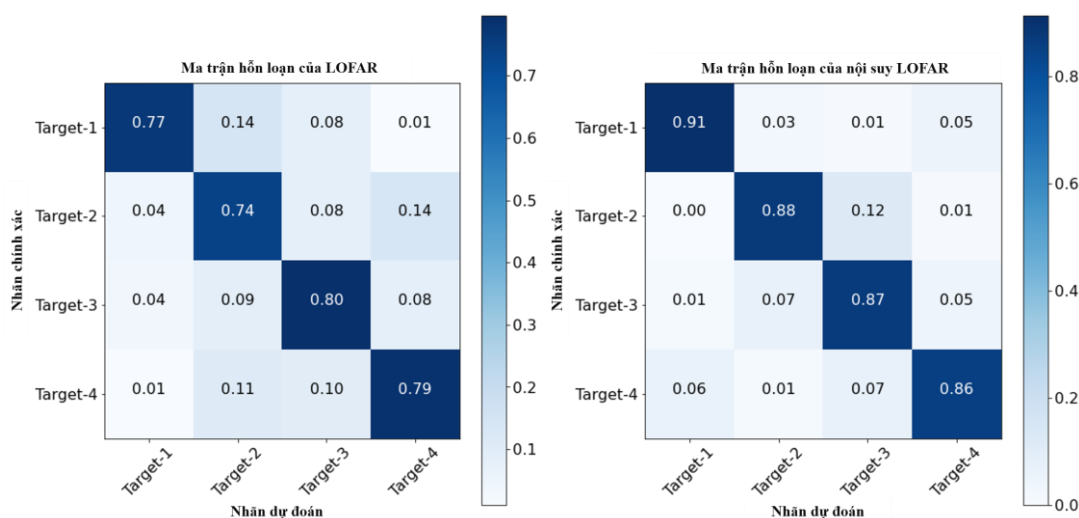


Hình 7. So sánh hàm mất mát của mô hình với hai bộ dữ liệu.

Kết quả phân loại chung giữa các nhóm đối tượng sau cải tiến đạt trung bình trên 85% so với trước đó là khoảng 77%, độ chính xác phân loại của từng lớp tàu được biểu diễn trong hình 8 thông qua một ma trận hỗn loạn.

Một trong những khó khăn lớn nhất trong xử lý tín hiệu thủy âm đó là hạn chế trong việc thu thập dữ liệu dưới nước vì các lí do dân sự và cả quân sự, đã dẫn đến việc thiếu dữ liệu và rất khó tiếp cận các dạng tín hiệu đặc thù. Trong thực tế, các dữ liệu thủy âm hiện nay tại Việt Nam thường thu thập trong các điều kiện môi trường, thời tiết khác nhau, các thiết bị thu cũng không đồng nhất, dẫn đến chất lượng tín hiệu gặp nhiều hạn chế. Các mô hình phân loại sẽ luôn đòi hỏi phải hoạt động hiệu quả với các bộ dữ liệu nhỏ, thiếu và tỷ lệ tín hiệu trên tạp thấp. Do đó, bài báo này đề xuất một giải pháp kết hợp giữa cải tiến kỹ thuật tiền xử lý LOFAR bằng phương pháp nội suy khối nối trực (cubic splines interpolation) và mô hình phân phối xác suất trong không gian ẩn (Variational autoencoder-VAE), để nâng cao khả năng phát hiện tín hiệu của chân

vật tàu trong các tập dữ liệu thực thu được (Bộ dữ liệu của các tàu hải quân thu được trong các quá trình triển khai hoạt động trên biển). Bằng cách cải tiến phương pháp trích xuất dữ liệu, sau đó ánh xạ dữ liệu từ không gian dữ liệu đầu vào sang không gian nén phi tuyến tính, thông qua các tổ hợp phi tuyến tính liên tục, mô hình đi sâu vào các cấu trúc dữ liệu phức tạp và biểu diễn các thuộc tính của tín hiệu trên một không gian đơn giản hơn và ít chiều hơn, do đó tăng mức độ trừu tượng của lớp đầu ra và cải thiện kết quả phân loại. Tất cả quá trình huấn luyện và thực thi mô phỏng được thực hiện trên máy tính trạm Dell T3600 Xeon 8 luồng, card đồ họa NVIDIA k2200 4GB trong môi trường Ubuntu 18.04, sử dụng CUDA10.1 và CuDNN7.6.5, với vòng lặp 200 lần.



**Hình 8.** Ma trận hỗn loạn biểu diễn độ chính xác phân loại cho từng lớp.

#### 4. KẾT LUẬN

Bài báo đã phân tích và cải tiến thuật toán LOFAR với đối tượng tàu sử dụng chân vịt làm giải pháp tiền xử lý dữ liệu trước khi đưa vào phân loại bằng mô hình xác suất phân phối trong không gian ẩn. Kết quả của mô hình đề xuất đã khắc phục được các hạn chế của việc thiếu hụt dữ liệu thủy âm, tăng độ chính xác của phép phân loại ngay cả khi SNR thấp. Tuy nhiên, để tiếp tục cải thiện tính chính xác, tốc độ xử lý, và từng bước cứng hóa các giải pháp, các hướng nghiên cứu sau này cần bổ sung ngưỡng phát hiện, kết hợp hai kỹ thuật DEMON và LOFAR cùng lúc, cũng như các giải pháp sinh dữ liệu mới như mạng GAN để tiếp tục cải thiện kết quả phân loại.

#### TÀI LIỆU THAM KHẢO

- [1]. W. S. Burdic, "Underwater Acoustic System Analysis", Peninsula Pub, (2003).
- [2]. R. O. Nielsen, "Sonar signal processing", Boston Artech House, pp. 16-85, (1991).
- [3]. J.C. Martino, "An approach to detect lofar lines", Pattern Recognition Letters 17.1, pp. 37-46, (1996).
- [4]. Q. Li, "Digital sonar design in underwater acoustics principles and applications", Springer Science & Business Media, (2012).
- [5]. J. Choi, "Acoustic Classification of Surface and Underwater Vessels in the Ocean Using Supervised Machine Learning", Sensors 19.16, (2019).
- [6]. Y. LeCun, "Deep learning", Nature, vol. 521, pp. 436-444, (2015).
- [7]. S. Min, "Underwater target recognition based on wavelet packet entropy and probabilistic neural network", International Conference on Signal Processing, Communication and Computing, IEEE, (2013).
- [8]. T. P. Hua, "Classification of Underwater Echo Based on Fractal Theory and Learning Vector Quantization Neural Network.", Applied Mechanics and Materials. Vol. 148. Trans Tech Publications Ltd, (2012).

- 
- [9]. Q. Weibiao, "Underwater targets classification using local wavelet acoustic pattern and Multi-Layer Perceptron neural network optimized by modified Whale Optimization Algorithm", Ocean Engineering 219: 108415, (2021).
- [10]. W. Zhengxian, et al "A method of underwater acoustic signal classification based on deep neural network", 2018 5th International Conference on Information Science and Control Engineering (ICISCE). IEEE, (2018).
- [11]. N. N. Moura, "Novelty detection in passive sonar systems using support vector machines", Latin America Congress on Computational Intelligence (LA-CCI), pp. 1-6, IEEE, (2015).
- [12]. T. McConaghy, H. Leung, and V. Varadan, "Classification of audio radar signals using radial basis function neural networks". IEEE Transactions on Instrumentation and Measurement, 52(6), pp. 1771-1779, (2003).
- [13]. M. Farrokhrooz and M. Karimi, "Ship noise classification using probabilistic neural network and AR model coefficients". Europe Oceans journal Vol. 2, IEEE, pp. 1107-1110, (2005).
- [14]. J. Q. Gauthier and T. A. Gooley, "Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians." Bone marrow transplantation 55.4, pp. 675-680, (2020).
- [15]. S. A Dyer and J. S. Dyer, "Cubic-spline interpolation", IEEE Instrumentation & Measurement Magazine 4.1, pp. 44-46, (2001).
- [16]. R. Fuji et al "Intention detection based on siamese neural network with triplet loss", IEEE Access 8: 82242-82254, (2020).
- [17]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, (2014).

#### ABSTRACT

#### **Classification of propeller vehicle using LOFAR cubic splines interpolation in combination with triple loss variational auto encoder**

*In the field of ocean acoustics, both traditional and modern underwater signal processing methods have recently achieved positive results. For sonar problems serving national defense and security tasks, the need for timely and accurate classification of propeller ship types is of top importance. This study presents an underwater signal processing model for the purpose of detecting and classifying propeller ships with improved LOFAR techniques by cubic splines interpolation (CSI) combined with probability distribution in the hidden space domain. The results of the proposed model, tested on real data sets, show that the classification accuracy has increased by 10%, achieving an efficiency of 88% compared to the previous models. This solution also demonstrates that the model combining traditional and modern methods can effectively classify actual signals even when the amount of data is lacking and the signal-to-noise ratio is low.*

**Keywords:** Underwater processing; Sonar; Interpolation; Triple loss.