Nâng cao hiệu năng nhận dạng tín hiệu thủy âm bằng mạng nơ-ron tích chập kết nối dư cải tiến

Đoàn Văn Sáng[1], Vi Công Đoàn[1], Trần Phú Ninh[1],
Nguyễn Văn Tiến[2], Trần Công Tráng[1*]

[1]Khoa Thông tin – Ra đa, Học viện Hải quân;
[2]Viện Tích hợp Hệ thống, Học viện Kỹ thuật quân sự.
[*]Email: trancongtrang@gmail.com
Nhận bài: 16/5/2022; Hoàn thiện: 22/6/2022; Chấp nhận đăng: 15/8/2022; Xuất bản: 26/8/2022.
DOI: https://doi.org/10.54939/1859-1043.j.mst.81.2022.53-59

TÓM TẮT

Bài báo trình bày kết quả nghiên cứu mô hình nhận dạng tín hiệu thủy âm sử dụng mạng nơ-ron tích chập theo cấu trúc kết nối dư được cải tiến từ mô hình ResNet (Residual Network) nhằm tăng hiệu năng về tốc độ xử lý mà vẫn đảm bảo độ chính xác nhận dạng cao. Khi so sánh với mô hình ResNet ban đầu và một số mô hình hiện có khác, mô hình đề xuất cho hiệu năng nhận dạng tốt về tỉ lệ nhận dạng đúng nguồn tín hiệu và tăng tốc độ dự đoán.

Từ khóa: Mạng nơ-ron nhân tạo; Mô hình ResNet; Phân loại tín hiệu thủy âm; Sôna thụ động.

1. ĐẶT VẤN ĐỀ

Phân loại tín hiệu thủy âm là một trong những nhiệm vụ đặc biệt quan trọng trong quân sự và nó cũng được sử dụng trong nhiều lĩnh vực dân sự. Trong các hoạt động dân sự, chẳng hạn như thăm dò biển, nhiệm vụ phân loại tín hiệu thủy âm giúp các nhà khoa học theo dõi, dự báo thủy văn và các hệ sinh thái biển dựa vào các đặc tính âm của từng loại sinh vật biển [1]. Ngày nay, lưu lượng tàu thuyền hoạt động trên biển ngày càng nhiều dẫn đến "ô nhiễm tiếng ồn" ảnh hưởng đến môi trường sinh thái của sinh vật biển. Việc thu và phân tích các tín hiệu âm trở lên phức tạp và phải diễn ra trong nhiều năm mới đưa ra được các giải pháp bảo vệ môi trường biển. Trong lĩnh vực quân sự, việc tự động phân loại tín hiệu thủy âm giúp trắc thủ nhanh chóng phát hiện và nhận dạng được mục tiêu, nâng cao hiệu quả trong tác chiến [2].

Gần đây, mạng nơ-ron nhân tạo là một trong những mô hình hữu ích ứng dụng các thuật toán trí tuệ nhân tạo (AI: Artificial Intelligence) để phân loại hình ảnh hoặc xử lý ngôn ngữ tự nhiên [3]. Hơn thế nữa, mạng nơ-ron nhân tạo cũng đã được ứng dụng nghiên cứu để nhận dạng giọng nói, phân loại loại âm thanh [4] và đạt được những kết quả nổi bật. Mặc dù có thể phức tạp hơn so với nhận dạng giọng nói, tín hiệu thủy âm cũng là một trong những dạng dữ liệu âm thanh có cùng tính chất nên nó cũng có thể nhận dạng được khi sử dụng mạng nơ-ron [2]. Chính vì vậy, áp dụng mạng nơ-ron nhân tạo để phân loại tín hiệu thủy âm, từ đó nhận dạng các nguồn phát xạ âm sẽ có tiềm năng để hỗ trợ trắc thủ ra quyết định nhận dạng mục tiêu. Điều này đã thúc đẩy nhóm tác giả xây dựng một mô hình mạng nơ-ron tích chập dựa theo cấu trúc của mô hình ResNet (Residual Network) nhưng đã cải tiến cho bài toán nhận dạng tín hiệu thủy âm nhằm đẩy nhanh tốc độ nhận dạng, nâng cao độ chính xác, và trợ giúp cho trắc thủ sôna thực hiện nhiệm vụ. Cụ thể, mô hình đề xuất đã được loại bỏ phần lớn các lớp chuẩn hóa và thay đổi kích thước kênh lọc trong các lớp tích chập nhằm giảm tải tính toán của mô hình, từ đó tăng tốc độ nhận dạng. Khi so sánh với mô hình ResNet ban đầu và một số mô hình hiện có khác trên cùng một tập dữ liệu gồm 12 loại tín hiệu thủy âm [5], mô hình ResNet cải tiến đã cho khả năng thực thi nhanh hơn mà vẫn đảm bảo độ chính xác nhận dạng tín hiệu thủy âm.

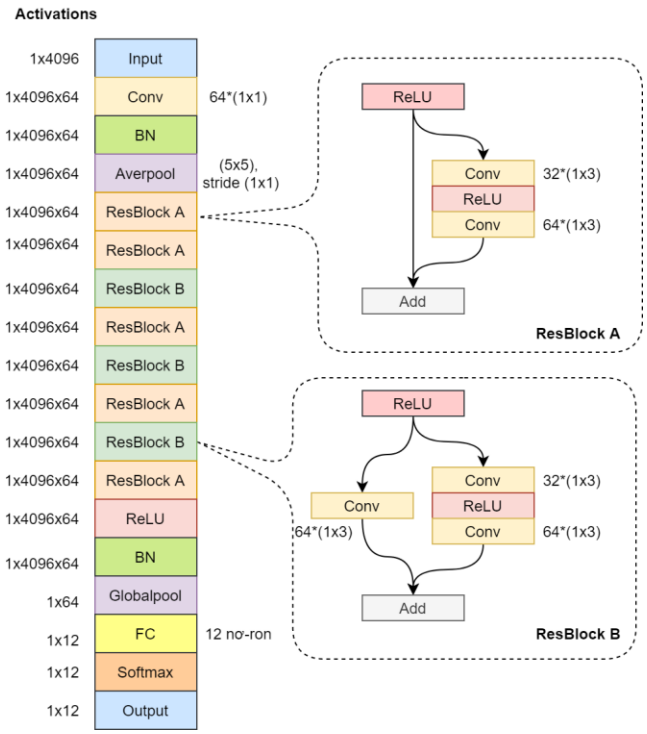
2. CHUẨN BỊ DỮ LIỆU CHO BÀI TOÁN

Để mô hình mạng nơ-ron nhân tạo có thể thực hiện được bài toán nhận dạng mục tiêu thủy âm cần phải có tập dữ liệu với độ tin cậy cao. Do đó, dữ liệu ShipEAR [5] được sử dụng trong

nghiên cứu này để huấn luyện cho mạng nơ-ron đề xuất, cũng như các mô hình hiện có khác để so sánh. Đây là tập dữ liệu được cấp quyền bởi Đại học Vigo, Tây Ban Nha. Tập dữ liệu gồm 96 tệp tin âm thanh của các loại tàu khác nhau với đầy đủ các thông tin về hình ảnh, tên tàu, kiểu loại tàu, tọa độ và tình huống tàu di chuyển. Sau khi nghiên cứu khai thác dữ liệu, nhóm tác giả đã tổng hợp thành 11 nhãn đại diện T01 đến T11 cho 11 nhóm dạng tiếng ồn chân vịt tàu và một dạng tiếng ồn tự nhiên. Để có tính tổng quát trong trường hợp không có bất kỳ nguồn phát nào trong môi trường biển thì vẫn tồn tại một dạng tiếng ồn, gọi là nhiễu tự nhiên trong môi trường biển. Vì vậy, một tệp nhiễu được giả lập để gộp cùng 11 tín hiệu nêu trên tạo thành một tập dữ liệu cho việc huấn luyện và thử nghiệm đối với mạng nơ-ron nhân tạo.

Mỗi tệp âm thanh được chia nhỏ thành nhiều đoạn tín hiệu với độ dài 4 096 mẫu, đảm bảo đủ dài để mô hình mạng nơ-ron có thể trích chọn được các đặc trưng hữu ích. Các mẫu âm thanh này được gán nhãn tương ứng với ký hiệu nhãn là Noise, T01 đến T11. Như vậy, để đảm bảo tính cân bằng cho dữ liệu, mỗi dạng âm thanh sẽ được lấy ngẫu nhiên 1 000 đoạn, mỗi đoạn có 4 096 mẫu. Để tăng thêm tính thử thách cho mạng nơ-ron, nhóm tác giả đã thêm các mức độ nhiễu Gauss khác nhau vào 11 tín hiệu gốc với các giá trị tỉ số tín/tạp (SNR: Signal to Noise Ratio) thay đổi từ -10 dB đến 20 dB, bước nhảy 2 dB vì nhiễu Gauss có phân bố chuẩn và tính chất gần với điều kiện thực tế. Như vậy, tổng thể có 192 000 đoạn âm thanh được gán nhãn tương ứng.

3. MÔ TẢ CẤU TRÚC MẠNG NƠ-RON ĐỀ XUẤT



Hình 1. *Sơ đồ cấu trúc của mạng nơ-ron đề xuất.*

Để có thể nhận dạng được các dạng tín hiệu thủy âm, một mạng nơ-ron nhân tạo được đề xuất trong nghiên cứu này. Theo đó, nhóm tác giả đã lựa chọn thiết kế mạng nơ-ron theo mô hình cấu trúc của mô hình ResNet [6] nhưng có sự cải tiến nhằm tăng tốc độ tính toán cho mô hình. Đây là mô hình được ứng dụng rộng rãi và có hiệu năng tốt đối với các bài toán phân loại. Dựa vào các "khối dư (Residual module)" và "kết nối bỏ qua (Skip-connection)", mà các đặc tính đại diện riêng cho từng dạng tín hiệu thủy âm sẽ được tự động trích chọn và cho kết quả nhận dạng với độ chính xác cao. Ngoài ra, việc sử dụng mô hình ResNet cải tiến cũng sẽ hạn chế được hiện tượng

nơ-ron bị tê liệt trong quá trình huấn luyện mạng cũng như giảm thiểu hiện tượng quá phù hợp (over-fitting). Cấu trúc của mô hình mạng đề xuất trong nghiên cứu này được thể hiện trong hình 1, trong đó có thể thấy, mạng gồm nhiều lớp được kết nối với nhau theo dạng khối dư và kết nối bỏ qua.

Các lớp được sử dụng gồm có: lớp đầu vào (Input), lớp tích chập (Conv: Convolution), lớp chuẩn hóa theo cụm (BN: Batch Normalization), lớp hàm kích hoạt (activation, ReLU: Rectified Linear Unit), lớp gộp trung bình (Averpool: Average pool), lớp kết nối đầy đủ (FC: Fully Connected), lớp Softmax và lớp đầu ra (Output). Như vậy, mạng ResNet cải tiến có tổng cộng 53 lớp, các lớp được mô tả như sau:

- Lớp đầu vào (Input) có kích thước là 4 096 phù hợp đoạn tín hiệu có độ dài 4096 mẫu.

- Các lớp tích chập (Conv) đóng vai trò như bộ lọc và chia kênh, được sử dụng để tự động tăng cường các đặc tính đại diện của từng dạng tín hiệu, đồng thời làm suy yếu những đặc tính gây nhiễu, không rõ nét hoặc không có tính phân biệt. Trong mạng nơ-ron này, có tổng cộng 20 lớp Conv được sử dụng. Thay vì sử dụng các lớp Conv với cửa sổ bộ lọc 2 chiều như mô hình ResNet gốc, mô hình cải tiến sử dụng cửa sổ một chiều với kích thức 1×3 để phù hợp với cấu trúc dữ liệu một chiều của tín hiệu âm thanh trong miền thời gian. Điều này giúp giảm kích thước của mô hình ResNet cải tiến và tăng tốc độ thực thi cho mô hình. Công thức tính tích chập một chiều được mô tả như sau [7]:

$$y_j = \sum_{k=-p}^{p} x_{j-k} w_k \,, \tag{1}$$

trong đó, x là chuỗi dữ liệu đầu vào, w là trọng số của kênh lọc và y là chuỗi dữ liệu đầu ra.

- Các lớp chuẩn hóa theo cụm (BN) được sử dụng như một phương pháp để chuẩn hóa dữ liệu, từ đó làm cho mạng nơ-ron được huấn luyện nhanh hơn và ổn định hơn. Trong mạng ResNet gốc, lớp BN được sử dụng theo sau mỗi lớp Conv, tuy nhiên, việc sử dụng quá nhiều lớp BN sẽ khiến cho mô hình phải sử dụng nhiều phép toán chuẩn hóa khi thực hiện quá trình nhận dạng. Để khắc phục vấn đề này, nhóm tác giả đã khéo léo loại bỏ các lớp BN này và thiết kế lại lớp BN sau lớp Conv đầu tiên và phía sau lớp ReLU cuối cùng. Như vậy, số lượng lớp BN được giảm đi đáng kể mà vẫn bảo đảm tốc độ và độ ổn định trong quá trình huấn luyện. Phép toán chuẩn hóa dữ liệu theo cụm dữ liệu được mô tả như sau [8]:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \,, \tag{2}$$

trong đó, x_i và $\hat{x}_i$ lần lượt là dữ liệu đầu vào và đầu ra của lớp batchnorm, μ_B và σ_B^2 lần lượt là giá trị trung bình và phương sai của cụm dữ liệu cho một lượt tính toán, ε là hằng số nhằm đảm bảo tính ổn định của phép toán khi phương sai quá nhỏ.

- Các lớp kích hoạt (activation, ReLU) sử dụng hàm kích hoạt ReLU để kích hoạt các đặc tính dương và đưa về "0" các đặc tính âm của dữ liệu. Có thể nói, mạng nơ-ron sẽ không thể huấn luyện nếu không có hàm kích hoạt. Có 17 lớp kích hoạt ReLU trong mạng nơ-ron. Hàm kích hoạt ReLU có thể được mô tả như sau [9]:

$$f(x) = \begin{cases} x \text{ khi } x > 0 \\ 0 \text{ khi } x \le 0 \end{cases} \tag{3}$$

- Lớp gộp trung bình (Averpool) dùng để tính giá trị trung bình cho mỗi lớp. Nó thực hiện việc lấy mẫu bằng cách chia dữ liệu thành các vùng nhỏ và tính trung bình mỗi vùng đó.

- Lớp gộp trung bình toàn cục (Globalpool) dùng để tính giá trị trung bình toàn cục. Việc lấy mẫu được thực hiện bằng cách tính giá trị trung bình của toàn bộ dữ liệu trên mỗi kênh lọc.

- Lớp kết nối đầy đủ (FC) thực hiện duỗi thẳng dữ liệu thành một véc-tơ, sau đó nhân với một ma trận trọng số. Đầu ra của lớp kết nối đầy đủ trong nghiên cứu này bằng số lượng tín hiệu thủy âm cần phân loại, tức là bằng 12.

- Lớp Softmax dùng hàm Softmax để tính xác suất cho từng phân lớp đầu ra của 12 nhãn tín hiệu, từ đó làm cơ sở để ra quyết định dự đoán mục tiêu. Hàm Softmax được mô tả như sau [10]:

$$\rho_i(z) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}. \quad (4)$$

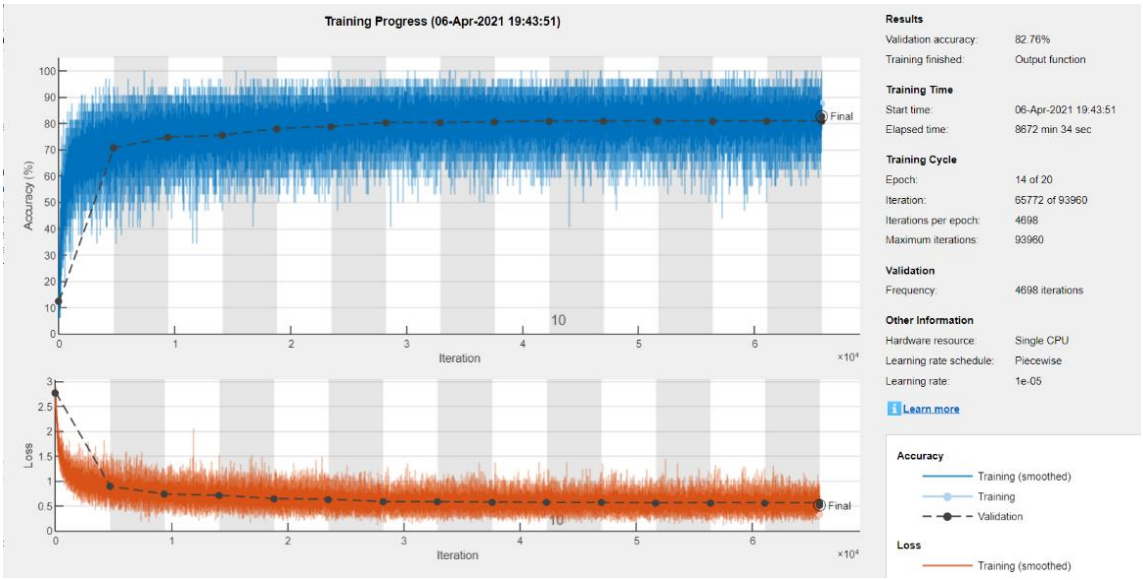
trong đó, z là dữ liệu đầu ra của lớp FC, $K = 12$ là số lượng phân lớp đầu ra.

- Lớp đầu ra (Output) được sử dụng để dự đoán phân lớp nguồn tín hiệu dựa vào giá trị xác suất tương ứng. Trong nghiên cứu này, lớp đầu ra sẽ quyết định bằng cách chọn nhãn có xác suất cao nhất, cụ thể như sau:

$$\text{Source}_{predicted} = \arg\max\{\rho(z)\} \quad (5)$$

4. ĐÁNH GIÁ HIỆU NĂNG NHẬN DẠNG TÍN HIỆU THỦY ÂM CỦA MẠNG NƠ-RON ĐỀ XUẤT

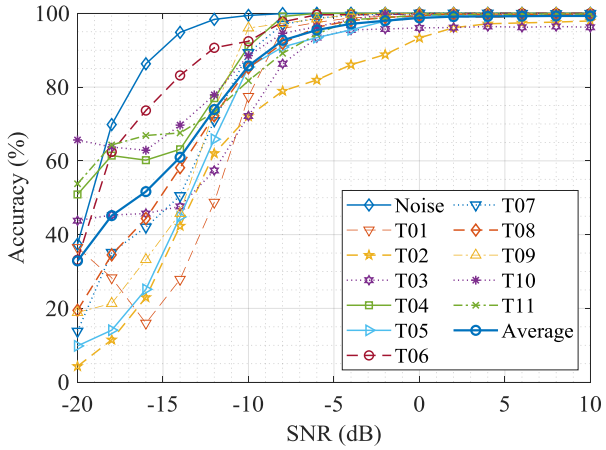
Mạng nơ-ron đề xuất được huấn luyện trên tập dữ liệu gồm 192 000 đoạn âm thanh có độ dài 4 096 mẫu với các giá trị SNR khác nhau từ -10 dB đến 20 dB. Quá trình huấn luyện trải qua 14 lần (epoch) với tổng thời gian là 8 672 phút (≈ 6,02 ngày), như thể hiện trong hình 2. Sau khi huấn luyện, độ chính xác nhận dạng tín hiệu thủy âm tương đối ổn định, đáp ứng được kỳ vọng của bài toán. Độ chính xác phân loại trung bình cho các mục tiêu và cho tất cả các giá trị SNR là 82,76%.



Hình 2. *Quá trình huấn luyện.*

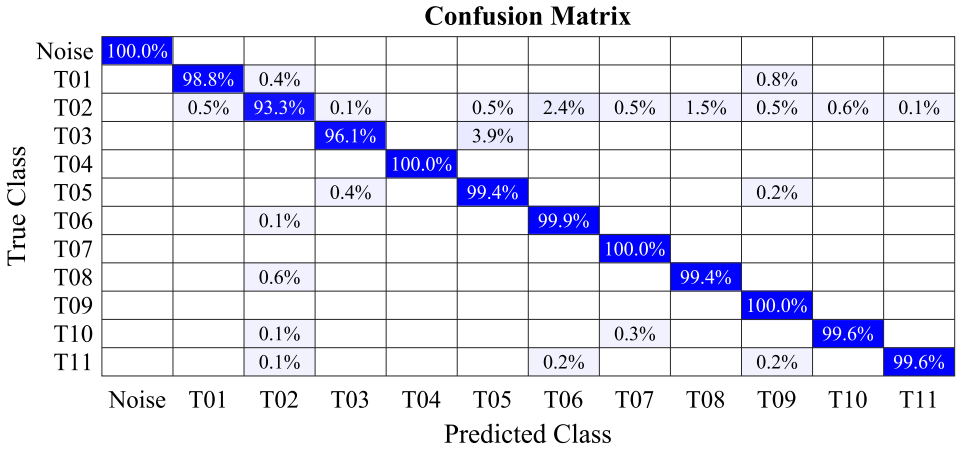
Tiếp theo, mô hình mạng nơ-ron đã huấn luyện được đánh giá khi nhận dạng tín hiệu thủy âm với các mức SNR khác nhau từ -10 dB đến 20 dB với cách bước 2 dB. Kết quả đánh giá thể hiện trong hình 3 cho thấy độ chính xác nhận dạng tăng lên khi tăng SNR, có nghĩa là, tín hiệu càng ít nhiễu thì chất lượng nhận dạng càng tốt. Cụ thể, tín hiệu nhiễu có chất lượng nhận dạng tốt hơn các tín hiệu còn lại với độ chính xác trên 99% khi SNR > -10 dB, vì nhiễu có đặc tính rất đặc thù nên mạng nơ-ron dễ dàng nhận biết đặc tính riêng so với các tín hiệu khác. Có thể thấy, tất cả các

tín hiệu đều đạt độ chính xác cao hơn 80% khi SNR > -5 dB. Độ chính xác trung bình đạt trên 80% khi SNR > -10 dB.



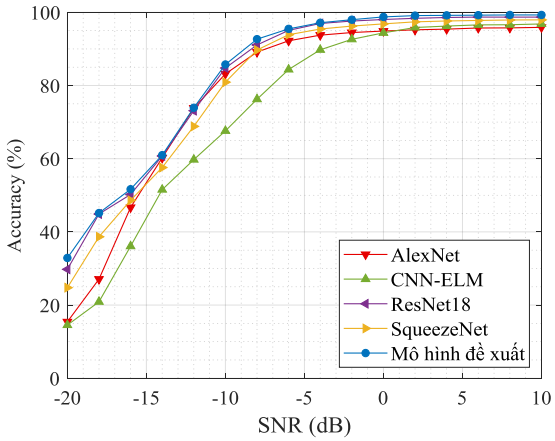
Hình 3. *Độ chính xác nhận dạng của từng tín hiệu thủy âm khi thay đổi SNR.*

Hình 4 thể hiện ma trận so sánh khi thực hiện nhận dạng tín hiệu thủy âm khi SNR = 0 dB. Kết quả cho thấy, mô hình mạng nơ-ron mà nhóm tác giả đề xuất đạt hiệu quả nhận dạng tốt với một số dạng tín hiệu như Noise, T04, T07 và T09 đạt 100% tỉ lệ nhận dạng đúng. Tín hiệu T02 có độ chính xác phân loại thấp nhất với 93,3%. Như vậy, có thể thấy rằng, mô hình mạng nơ-ron đề xuất có thể đáp ứng độ chính xác phân loại tốt, đạt 98,75% khi SNR = 0 dB.



Hình 4. *Ma trận so sánh nhận dạng tín hiệu thủy âm khi SNR = 0 dB của mạng nơ-ron nhân tạo đề xuất.*

Tiếp theo, mô hình đề xuất được so sánh với một số mô hình hiện có khác nhằm đánh giá những lợi thế mà nó mang lại. Theo đó, các mô hình gồm AlexNet [11], CNN-ELM [12], ResNet18 [6], SqueezeNet [13] và CNN-LSTM [14] được lựa chọn để so sánh với mô hình đề xuất. Các thông số so sánh gồm độ chính nhận dạng, số lượng trọng số và thời gian thực thi. Kết quả so sánh độ chính xác nhận dạng phụ thuộc vào SNR được thể hiện trong hình 5; có thể thấy rằng, mô hình đề xuất và ResNet18 có độ chính xác nhận dạng tín hiệu thủy âm gần tương đương nhau (mô hình đề xuất tốt hơn một chút so với ResNet18) và cả hai cho độ chính xác cao hơn các mô hình còn lại. Để đạt được độ chính xác như vậy là do mô hình đề xuất và ResNet18 đã sử dụng các kiểu kết nối bỏ qua và kết nối dư để thực hiện kết hợp giữa đặc tính hiện tại với đặc tính của lớp trước đó nhằm tránh việc bỏ sót các đặc tính hữu ích có tính phân biệt giữa các tín hiệu.



Hình 5. *So sánh độ chính xác phân loại của các mô hình cho cùng bài toán nhận dạng tín hiệu thủy âm.*

Kết quả so sánh số lượng trọng số của các mô hình, thời gian thực thi và độ chính xác trung bình được báo cáo trong bảng 1. Có thể thấy rằng, mô hình càng ít trọng số thì khả năng thực thi càng nhanh, do chúng thực hiện ít các phép toán trong quá trình thực hiện nhận dạng tín hiệu. Tuy nhiên, mặc dù mô hình CNN-LSTM có số lượng trọng số ít nhưng thời gian thực thi lại chậm nhất (khoảng 15 ms), điều này là do cấu trúc LSTM thực hiện các phép toán tuần tự nên kéo dài thời gian tính toán. Nhờ có thiết kế với ít trọng số nhất (137,2 nghìn) nên mô hình đề xuất đã đạt được thời gian thực thi ngắn nhất (3.5 ± 0,21 ms), và dựa vào sự kết hợp sơ đồ đặc trưng giữa các lớp trước và lớp sau mà mô hình đề xuất đã duy trì độ chính xác phân loại tương đương ResNet18 và cải thiện được độ chính xác trung bình cao hơn các mô hình được xem xét khác từ khoảng 3% (so với SqueezeNet) đến 10% (so với CNN-ELM). Các mô hình AlexNet, CNN-ELM và ResNet18 có số lượng trọng số lớn hơn nên thời gian xử lý bị chậm hơn.

Bảng 1. *So sánh số lượng trọng số, thời gian thực thi của các mô hình.*

Mô hình	Số lượng trọng số	Thời gian thực thi (ms)	Độ chính xác trung bình cho tất cả SNR
AlexNet	153,3 triệu	5,3 + 0,17	78,1%
CNN-ELM	41,2 triệu	4,7 ± 0,28	73,1%
ResNet18	11,1 triệu	4.39 ± 0,22	82,2%
SqueezeNet	727,5 nghìn	3,6 ± 0,25	80,0%
CNN-LSTM	161,0 nghìn	15 ± 0,24	79,5%
Đề xuất	137,2 nghìn	**3,5 ± 0,21**	**83,0%**

5. KẾT LUẬN

Như vậy, bài báo đã giải quyết bài toán nhận dạng tín hiệu thủy âm dựa vào mạng nơ-ron ResNet cải tiến. Mô hình mà nhóm tác giả đề xuất đã lược bỏ các lớp chuẩn hóa trong các khối kết nối bỏ qua và kết nối dư nhằm tăng tốc độ thực thi trong quá trình nhận dạng tín hiệu thủy âm. Mô hình đề xuất đã được huấn luyện và kiểm tra với các mức nhiễu khác nhau. Kết quả kiểm tra thể hiện nhiễu càng thấp thì chất lượng nhận dạng càng tốt, và đạt độ chính xác trung bình cao hơn 98,75% khi SNR ≥ 0 dB. Khi so sánh với mô hình gốc RestNet18, mô hình đề xuất cho độ chính xác tương đương nhưng thời gian thực thi nhanh hơn và số lượng trọng số ít hơn. Khi so sánh với một số mô hình hiện có khác, mô hình đề xuất đã đạt được hiệu năng cao hơn cả về độ chính xác nhận dạng, kích thước mô hình và thời gian thực thi.

TÀI LIỆU THAM KHẢO

[1]. K.J. Vigness-Raposa, G. Scowcroft, J.H. Miller, D. Ketten, *"Discovery of Sound in the Sea: An Online Resource,"* in Popper, A.N., Hawkins, A. (eds) The Effects of Noise on Aquatic Life. Advances in Experimental Medicine and Biology, vol 730. Springer, New York, NY, (2012), doi: 10.1007/978-1-4419-7311-5_30.

[2]. V. -S. Doan, T. Huynh-The and D. -S. Kim, *"Underwater Acoustic Target Classification Based on Dense Convolutional Neural Network,"* in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, Art no. 1500905, (2022), doi: 10.1109/LGRS.2020.3029584.

[3]. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, (2016).

[4]. A. B. Nassif, I. Shahin, I. Attili, M. Azzeh and K. Shaalan, *"Speech Recognition Using Deep Neural Networks: A Systematic Review,"* in IEEE Access, vol. 7, pp. 19143-19165, (2019), doi: 10.1109/ACCESS.2019.2896880.

[5]. D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, *"ShipsEar: An underwater vessel noise database,"* in Applied Acoustics, 113, 64-69, (2016).

[6]. K. He, X. Zhang, S. Ren, and J. Sun, *"Deep residual learning for image recognition,"* in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, pp. 770–778, Jun., (2016).

[7]. C. Lim, J. -Y. Kim and Y. Nam, *"ECG Signal Analysis for Patient with Metabolic Syndrome based on 1D-Convolution Neural Network,"* 2020 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 731-733, (2020).

[8]. Ioffe, Sergey, and Christian Szegedy. *"Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,"* [online] Available: https://arxiv.org/abs/1502.03167.

[9]. A. F. Agarap, *"Deep Learning using Rectified Linear Units (ReLU),"* [online] Available: https://arxiv.org/abs/1803.08375.

[10]. J. S. Bridle, *"Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters,"* in Proceedings of the 2nd International Conference on Neural Information Processing Systems (NIPS'89), MIT Press, Cambridge, MA, USA, pp. 211–217, (1989).

[11]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, *"ImageNet classification with deep convolutional neural networks,"* in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12), Curran Associates Inc., Red Hook, NY, USA, pp. 1097–1105, (2012).

[12]. G. Hu, K. Wang, Y. Peng, M. Qiu, J. Shi, and L. Liu, *"Deep learning methods for underwater target feature extraction and recognition,"* Comput. Intell. Neurosci., vol. 2018, pp. 1–10, Mar., (2018).

[13]. F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, *"SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5MB model size,"* (2016), arXiv:1602.07360. [Online]. Available: http://arxiv.org/abs/1602.07360.

[14]. X. C. Han, C. Ren, L. Wang, Y. Bai, *"Underwater acoustic target recognition method based on a joint neural network,"* in PLoS ONE 17(4), (2022), doi: 10.1371/journal.pone.0266425.

ABSTRACT

Improving the performance of underwater acoustic signal recognition using modified residual convolutional neural network

This paper presents the research results of an underwater acoustic signal recognition model using a convolutional neural network based on the residual structure, which is modified from the ResNet model to increase the performance in terms of processing speed while ensuring high recognition accuracy. Compared with the original ResNet model and some other existing models, the modified ResNet model provided a good recognition performance in terms of correct signal source recognition rate and increased prediction speed.

Keywords: Artificial neural network; ResNet model; Underwater acoustic signal classification; Passive sonar.