

Giải pháp nâng cao xác suất nhận dạng đúng mục tiêu UAV ứng dụng trí tuệ nhân tạo

Nguyễn Văn Trà^{1*}, Vũ Chí Thanh¹, Đoàn Văn Sáng²

¹Viện Ra đa, Viện Khoa học Công nghệ quân sự;

²Khoa Thông tin - Ra đa, Học viện Hải quân.

*Email: sinhviennaolin2000@gmail.com

Nhận bài: 12/9/2023; Hoàn thiện: 10/11/2023; Chấp nhận đăng: 15/11/2023; Xuất bản: 10/12/2023.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.FEE.2023.105-110>

TÓM TẮT

Trong bài báo này, chúng tôi đề xuất xây dựng một mô hình mạng nơ-ron có cấu trúc Residual-Inception kết hợp (đặt tên là RINet) và sử dụng hàm mất mát Focal Loss nhân trọng số trong quá trình huấn luyện để tiến hành nhận dạng mục tiêu ra đa dựa trong dataset RAD-DAR. Mô hình RINet kết hợp hàm mất mát đề xuất có độ chính xác nhận dạng mục tiêu trung bình là 98.72%, trong đó, xác suất nhận dạng đúng UAV lên đến 99.81%.

Từ khóa: UAV; Kết nối tắt; Hàm mất mát; Nhận dạng mục tiêu ra đa; Học sâu.

1. ĐẶT VẤN ĐỀ

Máy bay không người lái (UAV: Unmanned Aerial Vehicles) đang được sử dụng rộng rãi lĩnh vực dân sự và quân sự [1, 2]. Trong các cuộc chiến tranh hiện đại, việc sử dụng UAV đã làm thay đổi hoàn toàn cách thức tiến hành các cuộc xung đột vũ trang. Loại vũ khí này đảm trách các nhiệm vụ khó khăn hơn, gồm điều phối hành động của lực lượng mặt đất, điều chỉnh mục tiêu của pháo thủ trên máy bay có người lái và pháo binh, xác định hệ thống phòng không của đối phương và ném bom hoặc tấn công cảm tử các mục tiêu trên mặt đất. Vì vậy, việc phát triển các thiết bị có khả năng phát hiện và cảnh báo UAV là vấn đề mang tính cấp bách.

Các hệ thống ra đa đã được chứng minh là một giải pháp tốt cho việc phát hiện sớm mục tiêu UAV [3]. Trong đó, phân loại mục tiêu UAV với các mục tiêu khác là rất quan trọng [3-5] để có biện pháp đối phó với máy bay không người lái, chẳng hạn như bật hệ thống gây nhiễu, chế áp UAV hoặc kích hoạt các hệ thống đánh chặn.

Trong tác chiến thực tế, giá phải trả khi bỏ sót mục tiêu UAV (nhận dạng nhầm mục tiêu UAV thành loại mục tiêu khác) lớn hơn rất nhiều so với trường hợp khác. Vì vậy, bài báo này đề xuất giải pháp để làm tăng xác suất nhận dạng đúng đối với mục tiêu UAV. Chúng tôi đề xuất mô hình mạng nơ-ron RINet với hàm mất mát Focal Loss nhân trọng số để tiến hành huấn luyện và nhận dạng mục tiêu ra đa trên tập dữ liệu dataset RAD-DAR. Các chỉ số: độ chính xác, kích thước mô hình và thời gian tính toán được sử dụng để so sánh chất lượng của giải pháp đề xuất với các mô hình mạng nơ-ron trong công trình nghiên cứu khác đã công bố.

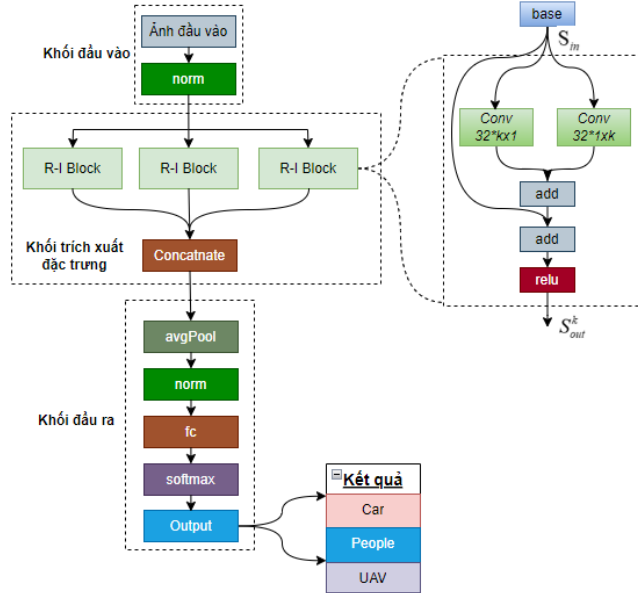
Nội dung tiếp theo của bài báo được trình bày theo bố cục sau. Phần 2 mô tả mô hình mạng nơ-ron RINet đề xuất kết hợp với hàm mất mát Focal Loss nhân trọng số. Sự so sánh chất lượng nhận dạng mục tiêu của mô hình đề xuất với các mô hình khác và đánh giá tính hiệu quả của việc cải tiến hàm mất mát được trình bày ở phần 3. Cuối cùng, phần 4 trình bày kết luận và hướng phát triển tiếp theo của bài báo.

2. ĐỀ XUẤT MÔ HÌNH MẠNG NƠ-RON

2.1. Mô hình RINet

Để tiến hành nhận dạng mục tiêu ra đa từ tập dữ liệu RAD-DAR, mạng nơ-ron RINet là sự kết hợp cấu trúc Inception [8] và ResNet [9] đã được xây dựng trong nghiên cứu này. Như mô tả trên hình 1, mô hình RINet bao gồm 3 khối chính: khối đầu vào, khối trích xuất đặc trưng và

khối đầu ra. Khối đầu vào có kích thước là 11x61 điểm ảnh tương ứng với kích thước ảnh của tập dữ liệu RAD-DAR. Ở tầng trích xuất đặc trưng, mô hình RINet bao gồm 3 khối Residual-Inception (R-I block) kết nối song song với nhau.



Hình 1. Mô hình mạng nơ-ron RINet đề xuất.

Khối R-I có cấu trúc là sự kết hợp giữa cấu trúc Inception và Residual. Mỗi mô đun inception trong khối R-I bao gồm 2 bộ tích chập song song, đầu vào và đầu ra của mô đun được thực hiện phép kết nối tắt theo cấu trúc Residual. Số bộ lọc của các bộ tích chập là 32, kích thước bộ lọc của 2 nhánh tích chập tương ứng là 1xk và kx1 (bộ lọc 1 chiều). Việc lựa chọn kích thước của các bộ lọc là 1xk và kx1 thay vì các bộ lọc kxk sẽ làm giảm số lượng tham số huấn luyện của mạng nơ-ron. Kích thước của bộ lọc (k) trong 3 khối R-I trong mô hình RINet này được thiết lập tương ứng là 3, 5, và 7. Việc sử dụng cấu trúc Inception và Residual và các bộ lọc có kích thước khác nhau trong khối tích chập giúp cải thiện độ chính xác phân loại của mạng nơ-ron so với các cấu trúc mạng nơ-ron khác [10].

Tại khối đầu ra, các phép tính avgPooling, Normalization được thực hiện trước khi hợp nhất (dense) với hàm kích hoạt softmax. Kích thước đầu ra lớp dense là 3 tương ứng với 3 loại mục tiêu mà mô hình cần nhận dạng.

Với tín hiệu đầu vào S_{in} cho lớp đầu vào, chúng ta có thể thu được đầu ra của khối R-I theo phương trình sau:

$$S_{out}^k = \delta(S_{in} + (F_{k \times 1}(S_{in}) + F_{1 \times k}(S_{in}))) \quad (1)$$

Trong đó, $F_{k \times 1}$ và $F_{1 \times k}$ là kết quả đầu ra của phép tích chập với kích thước của bộ lọc là kx1 và 1xk, σ là phép tính của hàm kích hoạt ReLU.

Hàm kích hoạt sử dụng trong các khối R-I là hàm ReLU (Rectified Linear Unit) như mô tả ở công thức (2) được sử dụng để kích hoạt đầu ra mỗi khối R-I.

$$Y = \text{Max}(0; x) \quad (2)$$

Dữ liệu đầu ra của 3 nhánh tích chập R-I Block sẽ được xếp chồng tại lớp (Concatenate) theo công thức 3 trước khi chuyển đến khối đầu ra tiến hành nhận dạng mục tiêu.

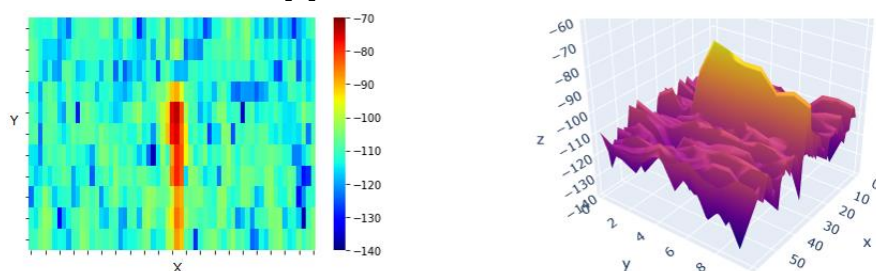
$$F_{concat} = [S_{out}^1, S_{out}^2, S_{out}^3] \quad (3)$$

Một số đặc trưng của mô hình RINet giúp nâng cao chất lượng nhận dạng mục tiêu như sau:

- Sử dụng các khối tính tích chập song song theo cấu trúc Inception sử dụng các bộ lọc kích thước của bộ lọc khác nhau sẽ làm tăng khả năng trích xuất được các đặc trưng từ ảnh dữ liệu vào.
- RINet sử dụng các bộ lọc có kích thước $1 \times k$ và $k \times 1$ thay vì các bộ lọc kích thước $k \times k$ trong mô hình sẽ làm giảm số lượng các tham số huấn luyện của mạng nơ-ron.
- Một ưu điểm khác của mô hình RINet là tránh được vấn đề không hội tụ (vanishing gradient problem) và overfitting thông qua việc sử dụng cấu trúc Residual.

2.2. Tập dữ liệu và thiết lập thử nghiệm

Hệ thống ra đa được sử dụng để thu thập tập dữ liệu RAD-DAR (Radar with Digital Array Receiver) là một ra đa phát tín hiệu FMCW băng X có 1 anten phát, 8 anten thu, tần số trung tâm 8.75 GHz, độ di tần 200 MHz [5].



Hình 2. Mẫu mục tiêu ra đa dạng heatmap 2D (ảnh bên trái) và 3D (ảnh bên phải).

Bộ dataset này bao gồm 17.485 ảnh mục tiêu trên miền cự ly – vận tốc Doppler được gán nhãn tương ứng 3 loại mục tiêu: người đi bộ, ô tô và UAV. Dữ liệu mẫu mỗi mục tiêu được lưu ở một file có định dạng .csv có kích thước 11 x 61 điểm ảnh (11 điểm ảnh trên trục cự ly, 61 điểm ảnh trên trục tần số Doppler). Giá trị tại mỗi điểm ảnh thể hiện công suất của tín hiệu thu được tại ô dữ liệu tương ứng với đơn vị đo là dBm. Cường độ tín hiệu phản xạ về của mục tiêu trong tập dữ liệu có giá trị nằm trong khoảng từ -140 dBm đến -70 dBm. Hình 2 thể hiện mẫu mục tiêu UAV trong bộ dữ liệu ở dạng heatmap 2D và 3D, với 3 trục x, y, z tương ứng thể hiện: tần số Doppler, cự ly và biên độ. Trong tổng số 17.485 mẫu dữ liệu, 6700 mẫu mục tiêu người đi bộ, 5720 mẫu mục tiêu ô tô và 5065 mẫu mục tiêu UAV tương ứng với tỷ lệ phần trăm là 38,32%, 32,71% và 28,97%.

2.3. Sử hàm mất mát Focal Loss nhân trọng số

Mạng nơ-ron được huấn luyện thông qua sử dụng các thuật toán tối ưu để điều chỉnh các hệ số trọng lượng theo một quy luật điều chỉnh nhất định (thường là phương pháp Gradient Descent). Là một phần của thuật toán tối ưu, giá trị lỗi của trạng thái mô hình hiện tại phải được tính toán một cách liên tục. Hàm mất mát (loss function) chính là một hàm toán học để tính giá trị lỗi, từ giá trị lỗi này các thuật toán tối ưu sẽ tiến hành cập nhật lại các trọng số mạng nơ-ron để đảm bảo giá trị lỗi đó giảm dần trong các lần đánh giá tiếp theo. Lựa chọn hàm mất mát phù hợp trong quá trình huấn luyện sẽ góp phần nâng cao hiệu năng cho mô hình mạng nơ-ron học sâu. Trong các bài toán phân loại mục tiêu, hàm mất mát Cross Entropy như mô tả ở công thức (4) thường được sử dụng để huấn luyện mạng nơ-ron.

$$CE(p_t) = -\log(p_t), \quad (4)$$

Trong đó, p_t là xác suất dự báo cho lớp đối tượng t . Trong Cross Entropy ta thấy rằng vai trò đóng góp vào loss function của các class cùng bằng.

Khi xảy ra hiện tượng mất cân bằng, chúng ta muốn rằng mô hình sẽ dự báo chuẩn hơn đối với những class thiểu số. Do đó, cần một hàm loss function hiệu quả hơn, có thể điều chỉnh được giá trị phạt lớn hơn nếu dự báo sai đối với nhóm thiểu số. Mục đích là để hạn chế dự báo sai

nhóm thiếu số vì nếu dự báo sai nhóm thiếu số thì hàm loss function sẽ trở nên lớn hơn. Tác giả Tsung-Yi và các cộng sự đã đề xuất hàm mất mát Focal Loss, đây là một sự điều chỉnh trong hàm Cross Entropy để giải quyết triệt để ảnh hưởng của mất cân bằng dữ liệu [6].

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) \quad (5)$$

Trong đó, $\alpha_t = \frac{1}{f_t + \varepsilon}$ là hệ số điều chỉnh để cân bằng hàm Cross Entropy với f_t là tần suất của đối tượng t , ε là một số dương rất nhỏ được cộng thêm để tránh phép chia 0, $(1-p_t)^\gamma$ là phần tử nhân thêm của hàm Focal Loss, γ thường được chọn trong khoảng từ 0 đến 5. Phần tử $(1-p_t)^\gamma$ trong hàm Focal Loss có tác dụng rất lớn trong việc điều chỉnh ảnh hưởng của nhãn lên đồng thời loss function và gradient descent.

Trong trường hợp của bài báo, tập dữ liệu RAD-DAR có số lượng 3 loại mục tiêu xe ô tô, người đi bộ và UAV tương đối cân bằng, chính vì vậy, thành phần α_t và $(1-p_t)^\gamma$ trong công thức (5) có tác động nhỏ đến giá trị hàm mất mát. Để cải thiện xác suất phát hiện đúng cho lớp mục tiêu UAV so với hai loại mục tiêu còn lại, chúng tôi đề xuất thay đổi sử dụng hàm mất mát Focal Loss nhân trọng số như mô tả ở công thức 6, ký hiệu là Weighted Focal Loss (WFL) để huấn luyện mô hình học sâu RINet đã đề xuất dùng.

$$WFL(p_t) = -\frac{w_t}{\sum_{t=1}^N w_t} (1-p_t)^\gamma \log(p_t) \quad (6)$$

Trong đó, hệ α_t được thay thế bằng giá trị ưu tiên thiết lập $\frac{w_t}{\sum_{t=1}^N w_t}$, với w_t hệ số ưu tiên của lớp

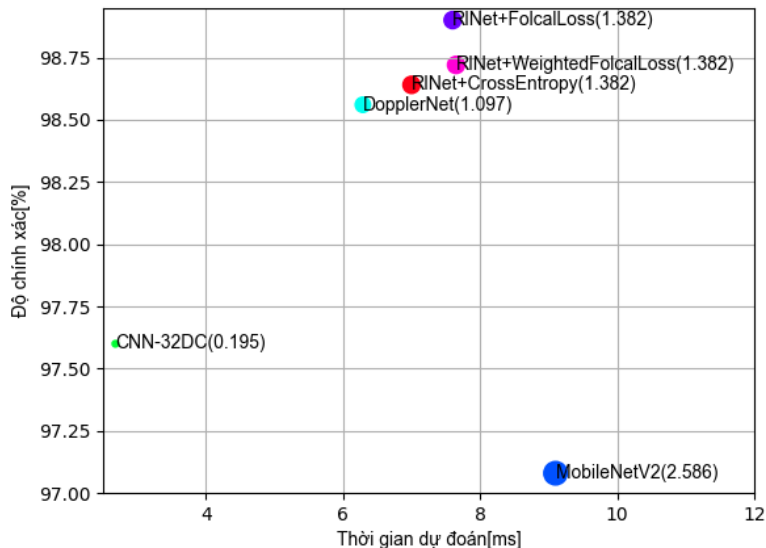
mục tiêu t . Bài báo xây dựng hàm mất mát WFL bằng ngôn ngữ lập trình Python, các giá trị w_t là các hệ số đầu vào có thể điều chỉnh. Cụ thể, bài báo thiết lập hệ số ưu tiên w_t của ba mục tiêu: người đi bộ, UAV và ô tô lần lượt là (1, 3, 1), tương ứng giá trị ưu tiên trong công thức (6) ba mục tiêu trên là (0.2, 0.6, 0.2). Hàm mất mát WFL bài báo đề xuất là sự điều chỉnh từ hàm mất mát Focal Loss để vừa giải quyết được vấn đề mất cân bằng dữ liệu vừa làm tăng xác suất nhận dạng đúng với lớp mục tiêu quan trọng so với các mục tiêu còn lại.

3. KẾT QUẢ VÀ THẢO LUẬN

Để đánh giá kết quả nhận dạng mục tiêu của mô hình mạng nơ-ron RINet đề xuất với hàm mất mát Focal Loss nhân trọng số, chúng tôi tiến hành so sánh mô hình đề xuất với mô hình DopplerNet, CNN-32DC và mô hình MobileNetV2. Mô hình DopplerNet và CNN-32DC là hai mô hình đã được xây dựng trong các công trình nghiên cứu [5] và [7] để tiến hành nhận dạng mục tiêu trên cùng tập dữ liệu RAD-DAR. Mô hình MobileNetV2 là một mô hình pretrained có cấu trúc residual rút gọn cho hiệu năng nhận dạng cao trong thư viện của framework Tensorflow. Các kết quả được tính trung bình của ba lần chạy huấn luyện mạng nơ-ron.

Hình 3 tổng hợp sự so sánh của mô hình đề xuất với các mô hình khác dựa trên các tiêu chí đánh giá: độ chính xác, kích thước mô hình và thời gian nhận dạng. Mô hình MobileNetV2 có kích thước mô hình lớn nhất (2.586 triệu tham số), thời gian dự đoán lớn nhất (9.1 ms) nhưng có độ chính xác nhận dạng thấp nhất (97.08%). Mô hình CNN-32DC có kích thước mô hình nhỏ nhất (0.195 triệu), tốc độ nhận dạng nhanh nhất (2.68 ms), tuy nhiên, độ chính xác chỉ đạt là 97.6% thấp hơn mô hình DopplerNet và RINet. Mô hình DopplerNet có kích thước nhỏ hơn và tốc độ nhận dạng nhanh hơn mô hình RINet đề xuất, thế nhưng độ chính xác nhận dạng của mô hình này nhỏ hơn các mô hình RINet. Ba mô hình RINet có kích thước như nhau (1.382 triệu

tham số) và thời gian nhận dạng mục tiêu gần như nhau. Trong đó, mô hình RINet với hàm mất mát Focal Loss đạt độ chính xác cao nhất (98.90%) so với mô hình RINet sử dụng hàm mất mát Cross Entropy (98.64%) và mô hình sử dụng hàm Focal Loss nhân trọng số (98.72%).



Hình 3. So sánh chất lượng của mô hình đề xuất với các mô hình khác.

Độ chính xác nhận dạng mục tiêu của mô hình RINet đề xuất với hàm mất mát Focal Loss nhân trọng số chỉ thấp hơn 0.018% so với mô hình RINet + Focal Loss. Tuy nhiên, có thể thấy trong bảng 1, mô hình RINet + Weighted Focal Loss với hệ số nhân trọng lượng của 3 loại mục tiêu (ô tô, UAV, người đi bộ) lần lượt là (0.2, 0.5, 0.2) cho xác suất nhận dạng mục tiêu UAV cao nhất tương ứng là 99.81%.

Bảng 1. So sánh độ chính xác nhận dạng của các mô hình mạng nơ-ron.

Mô hình mạng nơ-ron	Độ chính xác nhận dạng			
	Trung bình	Ô tô	Người đi bộ	UAV
DopplerNet + Cross Entropy	98.56%	98.22%	99.34%	98.13%
CNN-32DC + Cross Entropy	97.60%	97.57%	99.64%	95.63%
MobileNetV2 + Cross Entropy	97.08%	97.20%	98.36%	95.67%
RINet + Cross Entropy	98.64%	98.51%	99.06%	98.36%
RINet + Focal Loss	98.90%	98.62%	99.58%	98.52%
RINet + Weighted Focal Loss	98.72%	97.21%	99.15%	99.81%

Chúng tôi tiến hành huấn luyện các mô hình mạng nơ-ron đề xuất trên tập dữ liệu RAD-DAR với tỷ lệ 80% dữ liệu cho quá trình huấn luyện và 20% cho quá trình kiểm tra. Các tham số huấn luyện khác được thiết lập như sau: giá trị batch_size được thiết lập là 64, tốc độ học learning_rate ban đầu là 0.001 (giảm 90% sau mỗi 30 vòng lặp epochs). Thuật toán dừng huấn luyện sớm (Early Stopping) được áp dụng để kết thúc quá trình huấn luyện khi sau 5 chu kỳ chạy huấn luyện (epochs) mà không có sự cải thiện về độ chính xác của mô hình. Các mô hình mạng nơ-ron được xây dựng bằng ngôn ngữ lập trình Python, sử dụng framework Tensorflow. Các mô hình được huấn luyện và kiểm tra trên nền tảng phần cứng GPU của máy tính với cấu hình: chip xử lý Intel Xeon E5-2678 v3 CPU, card màn hình RTX 3060 GPU, Ram 32GB.

4. KẾT LUẬN

Bài báo đề xuất mô hình mạng nơ-ron RINet để tiến hành nhận dạng mục tiêu dựa vào dấu hiệu tần số Doppler của tập dữ liệu RAD-DAR. Mô hình RINet đề xuất có độ chính xác nhận dạng mục tiêu cao hơn so với các mô hình DopplerNet, CNN-32DC và MobileNetV2. Để nâng cao xác suất phát hiện đúng đối với mục tiêu là UAV so với các mục tiêu khác, bài báo đề xuất sử dụng hàm mất mát Focal Loss nhân trọng số. Kết quả là mô hình RINet sử dụng hàm mất mát Focal Loss nhân trọng số cho độ chính xác nhận dạng mục tiêu trung bình là 98.72%. Trong đó, mục tiêu UAV được nhận dạng chính xác với tỷ lệ 99.81%.

TÀI LIỆU THAM KHẢO

- [1]. Research, G.V. “*Commercial UAV Market Analysis By Product*”, (FixedWing, Rotary Blade, Nano, Hybrid), By Application (Agriculture, Energy, Government, Media and Entertainment) In addition, Segment Forecasts to 2022. 2016. Available online: <https://www.grandviewresearch.com/industry-analysis/commercialuavmarket> (accessed on 24 April 2019)
- [2]. X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi, J. Chen, “*Anti-drone system with multiple surveillance technologies: Architecture, implementation, and challenges*”, IEEE Commun. Mag. 56, 68–74, (2018).
- [3]. B. Torvik, K. E. Olsen, and H. Griffiths, “*Classification of birds and UAVs based on radar polarimetry*”, IEEE Geoscience and Remote Sensing Letters, vol.13, no.9, pp.1305-1309, (2016).
- [4]. P. Zhang, L. Yang, G. Chen and G. Li, “*Classification of drones based on micro-Doppler signatures with dual-band radar sensors*”, in Proc. 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL), Singapore, pp. 638–643, (2017).
- [5]. I. Roldan, C.R. del Blanco, A. Duque de Quevedo, F. Ibañez Urzaiz, J. Gismero Menoyo, A. Asensio López, D. Berjón, F. Jaureguizar, N. García, “*Dopplernet: a convolutional neural network for recognising targets in real scenarios using a persistent range–Doppler radar*”, IET Radar Sonar Navig. 14 (4), 593–600, (2020).
- [6]. Tsung-Yi Lin Pri “*Focal Loss for Dense Object Detection*” arXiv:1708.02002.
- [7]. Ann Janeth Garcia, Ali Aouto, Jae-Min Lee, Dong-Seong Kim “*CNN-32DC: An improved radar-based Drone recognition system based on Convolutional Neural Network*”, ScienceDirect accepted (2022).
- [8]. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. “*Going Deeper With Convolutions*” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-9, (2015).
- [9]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun “*Deep Residual Learning for Image Recognition*” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, (2016).
- [10]. Muhammad, W.; Bhutto, Z.; Ansari, A.; Memon, M.L.; Kumar, R.; Hussain, A.; Shah, S.A.R.; haheem, I.; Ali, S. “*Multi-Path Deep CNN with Residual Inception Network for Single Image Super-Resolution*”. Electronics, 10, 1979, (2021). <https://doi.org/10.3390/electronics10161979>

ABSTRACT

Solutions to increase the probability of the accuracy of UAV recognition target based on artificial intelligence

In this article, we propose to build a neural network model with a combined Residual-Inception structure (named RINet) and use the weighted Focal Loss loss during training to perform radar target classification based on the RAD-DAR dataset. The RINet model incorporating the proposed loss function has an average target classification accuracy of 98.72%, in which the probability of correctly identifying the UAV is up to 99.81%.

Keywords: UAV; Residual; Loss function; Radar target recognition; Deep learning.