«علوم و فناوری‌های پدافند نوین»
سال چهاردهم، شماره ۱، بهار۱۴۰۲؛ ص ۵۰-۳۹

علمی - پژوهشی

شناسایی صداهای منتشر شده از شناورهای سطحی با استفاده از الگوریتم کانولوشنی موبایل نت

حسن اکبریان [۱*]، محمدحسین صداقی [۲]

۱- دانشجوی دکتری، ۲- استاد، دانشگاه صنعتی سهند تبریز،ایران

(دریافت: ۱۴۰۱/۱۰/۲۱، بازنگری: ۱۴۰۲/۰۱/۲۲، پذیرش: ۱۴۰۲/۰۲/۲۲، انتشار: ۱۴۰۲/۰۳/۰۱)
DOR: https://dor.isc.ac/dor/20.1001.1.26762935.1402.14.1.4.4

چکیده

با حرکت شناورها بر روی آب و فعالیت موتورهای پیشرانه و چرخش پروانه‌های آن، سیگنال‌های صوتی از آنها منتشر می‌شود که اصطلاحاً به آن نویز منتشر شده از کشتی گفته می‌شود. امروزه نیروهای دریایی جهان با استفاده از این صداها، نسبت به شناسایی شناورهای سطحی عبوری از آب‌های سرزمینی و بین‌المللی اقدام می‌کنند. یکی از بهترین روش‌ها برای دسته‌بندی و شناسایی شناورها با توجه به صداهای منتشر شده از آنها، یادگیری عمیق است. با استفاده از یادگیری عمیق، ویژگی‌های منحصربه‌فرد سیگنال قابل‌استخراج بوده که از دقت بالایی در شناسایی اهداف برخوردار است. در این مقاله مدلی مبتنی بر شبکه موبایل نت طراحی گردیده است که سیگنال‌های صوتی دریافت شده توسط گیرنده‌های صوتی زیر آب (هایدروفون‌ها) را پردازش نموده و درنهایت بادقت بالایی طبقه‌بندی می‌نماید. ورودی این مدل تصاویر طیف‌نگار مربوط به داده‌های صوتی سونار غیرفعال است که با استفاده از تبدیل فرکانسی کوتاه‌مدت (STFT) تولید شده‌اند. این مدل در برنامه پایتون و با استفاده از کتابخانه کراس ایجاد شده و نتایج به دست آمده نشان می‌دهد دقت شناسایی مدل پیشنهادی بیش از ۹۶٪ و زیان ارزیابی آن کمتر از ۳٪ است. نسبت به روش‌های متداول یادگیری عمیق، روش پیشنهادی علاوه بر داشتن سرعت محاسباتی مناسب، از دقت شناسایی قابل قبولی نیز برخوردار است.

کلیدواژه‌ها: شناورهای سطحی، یادگیری عمیق، سونار غیرفعال، شناسایی صوت زیر آب.

Recognition of Acoustic Emitted from Surface Vessels Using MobileNet Convolutional Algorithm

H. Akbarian [*], M. H. Sedaaghi
Sahand University of technology, Tabriz, Iran

(Received: 2023/01/11 ; revised: 2023/04/11 ; Accepted: 2023/05/12 ; published: 2023/05/22)

Abstract

With the movement of the vessels in the water and the activity of the propulsion engines and the rotation of its propellers, they emit sound signals from them, which are called the ship's radiated noises. Today, the naval forces of the world use these sounds to identify surface vessels passing through territorial and international waters. One of the best methods for classifying and recognizing vessels according to the sounds emitted by them is deep learning. By using deep learning, it can extract the unique features of the signal, which have high accuracy in recognition. This paper designed a model based on the Mobilenet network, which processes the acoustic signals received by underwater sound receivers (hydrophones) and finally classifies them with high accuracy. The input of this model is the spectrogram images related to passive sonar sound data, which are produced using short-term frequency transformation (STFT). We created this model in the Python program using the keras library and the results show that the accuracy of the proposed model is more than 96% and its evaluation loss is less than 3%. Compared to the common methods of deep learning, the proposed method, in addition to having a suitable calculation speed, also has an acceptable recognition accuracy.

Keywords: Surface Vessels, Deep Learning, Passive Sonar, Underwater Acoustic Recognition.

[*] Corresponding Author E-Mail:H_Akbarian98@Sut.ac.ir

۱. مقدمه

وقتی کشتی‌ها در آب حرکت می‌کنند، صداهایی تولید می‌کنند که اصطلاحاً به آنها صوت منتشرشده از کشتی[1] گفته می‌شود. با توجه به ویژگی‌های منحصربه‌فرد سیگنال‌های منتشر شده از کلاس‌های مختلف شناورهای سطحی و زیرسطحی، می‌توان با تجزیه‌وتحلیل این سیگنال‌های صوتی، کلاس خاصی از کشتی‌ها یا زیردریایی‌ها و یا حتی یک شناور ویژه را شناسایی کرد. تشخیص شناورهای عبوری از آبراه‌ها با استفاده از صوت زیر آبی منتشر شده از آن‌ها، یکی از مهم‌ترین و چالش‌برانگیزترین موارد در پردازش سیگنال صوتی زیر آب است [۱].

در روش‌های سنتی، شناسایی اهداف شناور زیر آب از طریق کاربران آموزش‌دیده سونار انجام می‌گیرد که به دلیل نیاز به پایش مداوم صوت‌های دریافتی توسط سونار غیرفعال و اثرات ناشی از شرایط جوّی می‌تواند نادرست باشد. ازاین رو، ایجاد سیستم شناسایی خودکار و قوی، به‌منظور جایگزینی برای کاربر انسانی ضروری به نظر می‌رسد. تشخیص هدف صوتی زیر آب، یک مسئله پیچیده‌ای از تشخیص الگو می‌باشد. به دلیل نوع مکانیزم تولید صوت‌های منتشر شده از شناورها و تأثیرپذیری آن‌ها از کانال‌های صوتی زیر آب، سیگنال تولید شده دارای خصوصیات نوسانی[2]، غیرایستا[3] بوده و غیر خطی می‌باشد [۲]. از مهم‌ترین منابع تولید صدا از سوی شناورها، سیستم پیشرانه شناور، پروانهٔ شناور و نویزهای هیدرودینامیکی است. سیستم پیشران شامل پره‌های بزرگ، چرخ‌دنده، موتور پیشران، توربین‌ها و... می‌باشد. صداهای ناشی از پروانهٔ شناور مهم‌ترین انواع نویز در بین صداهای ناشی از حرکت شناورها است. دریافت صداهای منتشر شده از شناورها در زیر آب توسط دستگاهی به نام هایدروفون[4] انجام می‌گیرد. هایدروفون‌ها نوعی گیرنده صوتی غیرفعال می‌باشند که با دریافت فشار حاصل از موج صوتی، آن را به سیگنال قابل‌پردازش خروجی تبدیل می‌نمایند. سونار غیرفعال سونواری است که تنها از گیرنده استفاده می‌کند. در این حالت سامانه با دریافت صوت تولیدی توسط اهداف و انجام پردازش بر روی آن، نوع شناورها را شناسایی می‌نماید [۳].

در راستای حفظ امنیت دریایی و جلوگیری از نفوذ دشمن، به‌منظور تشخیص صحیح و به‌موقع اهداف شناور سطحی، استفاده از روش‌های هوش مصنوعی، یادگیری ماشین و بالاخص روش نوین یادگیری عمیق، در شناسایی دقیق موقعیت و نوع اهداف، بسیار حائز اهمیت و ضروری است. به دلیل مزیت‌های ذاتی تشخیص اهداف صوتی به‌صورت غیرفعال، در مقایسه با روش‌های دیگر، این روش توسط نیروهای دریایی جهان مورد استفاده قرار می‌گیرد. روش پردازش کلی برای تشخیص صداهای منتشر شده از کشتی مبتنی بر داده‌های دریافتی از سونار غیرفعال، به این صورت است که ابتدا پیش‌پردازش لازم (شامل پنجره‌گذاری، فیلترینگ، حذف نویز، تعیین نرخ نمونه‌برداری و...) بر روی سیگنال صوتی اجرا می‌شود و با استفاده از تبدیل فرکانسی STFT طیف‌نگار دیتاست تولید شده به دست خواهد آمد. در مرحله استخراج ویژگی، خصوصیات خاصی از داده‌های سونواری برای کاهش هشدارهای کاذب[5] و افزایش نرخ شناسایی استخراج می‌شوند. مرحله بعدی، استفاده از ویژگی‌های استخراج شده به‌عنوان ورودی طبقه‌بندها است که در آن بخشی از داده‌ها برای آموزش و بقیه برای آزمایش و ارزیابی عملکرد مدل مورد استفاده قرار می‌گیرد. با توجه به عملکرد برجسته شبکه‌های عصبی عمیق (به‌ویژه روش‌های یادگیری عمیق) در تشخیص صدا، در این تحقیق از شبکه‌های یادگیری عمیق برای استخراج و تشخیص ویژگی‌های سیگنال صوتی کشتی در زیر آب استفاده شده است. یکی از روش‌های مدرن یادگیری عمیق در شناسایی اهداف، استفاده از شبکه کانولوشنی موبایل نت[6] است که در این مقاله از آن استفاده شده است [۴]. با انجام پیش‌پردازش‌های لازم بر روی مجموعه داده صوتی و استخراج ویژگی‌های زمان - فرکانس در قالب تصاویر طیف‌نگار، تصاویر مربوط به هر کلاس، به‌عنوان ورودی مدل پیشنهادی، به سیستم تزریق شده و نتایج حاصل بررسی گردیده است. با استفاده از این مدل پیشنهادی، می‌توان به‌دقت[7] شناسایی ٪۹۶/۲۷ و زیان ارزیابی[8] کمتر از ٪۳ دست‌یافت که نسبت به روش‌های دیگر شناسایی اهداف صوتی زیر آب[9]، از توانایی بیشتری برخوردار است.

در این مقاله پس از بیان مقدمه و بررسی مدل‌های ساختاری پیشین، روش پیشنهادی به تفصیل بیان می‌گردد. سپس مجموعه داده که شامل صدای شناورهای سطحی است، شرح داده شده و پیش‌پردازش و شبیه‌سازی بر روی این مجموعه داده اعمال می‌گردد. در نهایت نتایج حاصله با روش‌های معمول و جاری یادگیری عمیق برای شناسایی و طبقه‌بندی اهداف صوتی زیر آب مقایسه شده و نتیجه‌گیری بیان می‌گردد.

۱-۱. بررسی مدل‌های ساختاری پیشین

اولین مطالعه‌ای که از شبکه های عصبی برای تشخیص سیگنال سونار استفاده کرد توسط گورمن و شاینوسکی [۵] پیشنهاد شد. آنها از یک شبکه سه‌لایه (با یک‌لایه پنهان) برای طبقه‌بندی سیگنال های بازگشتی سونار فعال به دو کلاس سیلندر زیر آب و

[1] Ship-Radiated Sound
[2] Oscillation
[3] Non-Stationary
[4] Hydrophon

[5] False Alarm
[6] MobileNet
[7] Accuracy
[8] Loss Validation
[9] Underwater Acoustic Target Recognition

صخره استفاده کردند که نتایج طبقه بندی تا ۹۰/۴ درصد دقت را نشان داد. چن و همکاران [۶] از یک روش طبقه‌بندی بر اساس شبکه‌های عصبی برای شناسایی کشتی‌ها پیشنهاد کردند. آنها از تبدیل موجک برای استخراج ویژگی‌های تونال از داده‌های ورودی استفاده کردند و سپس از پرسپترون چندلایه برای طبقه بندی داده‌های پردازش شده استفاده کردند و توانستند به‌دقت ۹۴ درصد شناسایی دست یابند. دوبک و همکاران [۷] از شبکه های عصبی مبتنی بر روش نزدیک‌ترین همسایه k و طبقه‌بندی‌کننده فیلتر تشخیص بهینه برای شناسایی و طبقه‌بندی مین‌های دریایی استفاده کردند. آنها در این روش به دقت ۹۲/۶۴ درصد دست یافتند. باران و همکاران [۸] روشی را برای تشخیص سیگنال های دریافتی در سونار غیرفعال بر اساس شبکه‌های عصبی پیشنهاد کردند. آنها سیگنال های دریافتی را با استفاده از شبکه های عصبی به سه کلاس کشتی طبقه بندی کردند و دقت آن بیش از ۹۰ درصد بود. ویلیامز [۹] و گالوشا و همکاران [۱۰] از شبکه های عصبی کانولوشن برای طبقه بندی تصاویر به دست آمده در سونار روزنه مصنوعی (SAS) استفاده کردند. باخ و همکاران در [۱۱] تصاویر طیف‌نگار به دست آمده از مجموعه داده‌های صوتی با اندازه (۳×۲۲۴×۲۲۴) را به‌عنوان ورودی به شبکه‌های LeNet، VGG و CNN تغذیه نمودند. در این مدل از ۱۰۰ دوره برای آموزش انتخاب شده است. از تجزیه‌وتحلیل نتایج بین مدل‌ها، دقت LeNet تنها ۷۰ درصد بوده و دقت VGG ۷۸ و CNN به سطح دقت ۸۷ درصد رسیده است. زمان آموزش LeNet و CNN تقریباً ۱۵۰ دقیقه است، درحالی‌که برای VGG این زمان به ۱۸۰ دقیقه می‌رسد. یانگ و همکاران [۱۲] با استفاده از شبکه باور عمیق و ماشین بولتزمن محدود شده برای مقداردهی اولیه پارامتر، مدلی برای تشخیص اهداف صوتی زیر آب به‌وسیله یادگیری ویژگی‌های به دست آمده از اطلاعات تمیزدهنده اضافی مربوط به اهداف دارای برچسب و بدون برچسب، ارائه دادند. نتایج نشان داد که این سیستم به یک دقت طبقه بندی ۹۰/۸۹ درصد دست‌یافته است. ژیهان پارک [۱۳] در مدل پیشنهادی خود با استفاده از تخمین طیف فرکانسی باند باریک سیگنال های صوتی، تصاویر طیفی داده‌های صوتی سونار را به دست آورده و پس از تقسیم‌بندی به عنوان ورودی یک شبکه عصبی کانولوشنی ۱۱ لایه در نظر گرفته است تا پیش‌بینی نماید تکه‌های تصاویر پردازش شده بخشی از یک فرکانس صوتی می باشند و یا خیر. وی در ۱۰ لایه اول از هسته و در لایه آخر از هسته استفاده نموده است. با استفاده از این ساختار، ۹۲/۵ درصد دقت و ۹۹/۸ درصد صحت بازیابی به دست آمد. زنگ و همکاران [۱۴] با استفاده از ویژگی‌های مل اسپکتروگرام سه‌بعدی و شبکه ResNet اصلاح شده توانستند به‌دقت ۹۵/۵ درصد دست یابند. میائو و همکاران [۱۵] مدلی را پیشنهاد کردند که با استفاده از تبدیل فرکانسی_زمانی ویژگی‌های سیگنال صوتی را استخراج و پس از تغذیه آن به شبکه یادگیری عمیق CNN، کلاس‌های مختلف آن را طبقه‌بندی نموده است. همچنین یانگ و همکاران [۱۶] با تجزیه سیگنال‌های صوتی مربوط به ۵ کلاس از کشتی‌ها در غشای پایه توسط لایه‌های کانولوشن عمیق به همراه یک‌لایه تبدیل فرکانس زمانی، ویژگی های شنوایی را به‌صورت نظارت شده استخراج و با حداکثر دقت ۹۴/۷۵ درصد موفق به شناسایی اهداف شده است. زنگ و همکاران [۱۷] با ترکیب شبکه‌های عصبی ResNet و DensNet به مدلی دست یافته‌اند که دارای دقت شناسایی ۹۵/۶۱ درصد است. جیانگ، ژائو و وانگ [۱۸] مدلی با ترکیب شبکه خصمانه مولد کانولوشنال عمیق DCGAN اصلاح شده و مدل S-ResNet را برای به دست آوردن دقت طبقه بندی خوب پیشنهاد کردند که توانسته است به‌دقت طبقه‌بندی ۹۳/۰۴ درصد دست یابد. تیان، چن، وانگ و لیو [۱۹] یک واحد باقیمانده چند مقیاسی (MSRU) پیشنهاد کرده است که قادر به ساخت شبکه پشته کانولوشن عمیق است. این ساختار انعطاف پذیر و متعادل که شبکه عصبی عمیق باقیمانده چند مقیاسی (MSRDN) نامیده می‌شود، برای طبقه‌بندی هدف صوتی زیر آب بکار رفته و توانسته است به دقت شناسایی ۸۳/۱۵ درصد دست یابد. هو، وانگ و لیو [۲۰] یک شبکه عصبی کانولوشنال قابل تفکیک جدید برای شناسایی سیگنال‌های منتشر شده از کشتی‌ها در شکل موج‌های حوزه زمان پیشنهاد کرده است. ویژگی‌های عمیق حاوی اطلاعات داخلی هدف می باشند که توسط یک شبکه کانولوشن DWS استخراج می شوند. میانگین نرخ شناسایی طبقه بندی هنگام آزمایش بر روی یک مجموعه سیگنال آکوستیکی به۹۰/۹ درصد می‌رسد. چن و همکاران [۲۱] روشی برای شناسایی اهداف صوتی زیرآبی پیشنهاد کردند که طیف حاصل از ضبط آنالیز فرکانس پایین را به‌عنوان ورودی شبکه عصبی کانولوشنال (CNN) در نظر گرفته و یک CNN مبتنی بر LOFAR برای شناسایی آنلاین ایجاد کردند. روش LOFAR-CNN پیشنهادی توانسته است به دقت شناسایی ۹۵/۲۲ درصد دست یابد. صفری، ظهیری و خوزین قند [۲۲] با استفاده از مدولاسیون پروانه بر روی سیگنال ارسالی سونار میکرو داپلر و هسته های مختلف، یک ماشین بردار پشتیبان (SVM) برای تشخیص خودکار اهداف متحرک سوناری را پیشنهاد کرده‌اند. دقت شناسایی صحیح اهداف برای نسبت‌های مختلف سیگنال به نویز (SNR) و زوایای دید متفاوت بود. هانگ و دیگران [۲۳] یک روش طبقه‌بندی با استفاده از ویژگی‌های همجوشی و شبکه باقیمانده ۱۸ لایه (ResNet18) پیشنهاد کرده است. آزمایش های شناسایی بر روی مجموعه نویز منتشر شده از کشتی به نام ShipsEar از یک محیط واقعی انجام شده که دقت شناسایی ۹۴/۳ درصد را نشان می دهد. لی، سونگ و فنگ [۲۴] با ترکیب روش های تشخیص پوش مدولاسیون بر روی نویز (DEMON) و ضبط آنالیز فرکانس پایین (LOFAR) به‌منظور استخراج ویژگی های اهداف صوتی زیر آب و استفاده از شبکه عصبی

کانولوشنی CNN توانسته‌اند به‌دقت شناسایی ۹۴/۰۰ درصد دست یابند. کمال، چاندران و سوپریا [۲۵] با استفاده از روش حافظه طولانی کوتاه‌مدت[1] توانستند به‌دقت طبقه‌بندی ۹۵/۲ درصد در شناسایی اهداف سونار غیرفعال دست یابند.

۱-۲. یادگیری عمیق

یادگیری عمیق[2] زیرمجموعه ای از یادگیری ماشین[3] است که در آن از الگوریتم‌هایی استفاده می‌شود که مغز انسان را شبیه‌سازی می‌کنند. این الگوریتم‌ها شبکه‌های عصبی مصنوعی[4] نام دارند. یادگیری عمیق روشی از یادگیری ماشین است که به رایانه‌ها یاد می‌دهد کاری را که معمولاً انسان‌ها انجام می‌دهند، با مثال‌هایی یاد بگیرند. در یادگیری عمیق، یک مدل رایانه‌ای یاد می‌گیرد که روش‌های شناسایی و دسته‌بندی را مستقیماً بر روی تصویر، متن یا صدا اجرا نماید. مدل‌های یادگیری عمیق می‌توانند به بالاترین سطح دقت برسند، به‌نحوی‌که گاهی از انسان‌ها هم بهتر عمل می‌کنند. مدل‌های یادگیری عمیق به‌وسیله گروه‌های بزرگی از داده‌ها و شبکه‌های عصبی با لایه‌های بسیار، آموزش داده می‌شوند. شبکه های عصبی سنتی فقط شامل ۲ یا ۳ لایه پنهان هستند، درصورتی‌که شبکه‌های عمیق[5] می‌توانند بیش از ۱۵۰ لایه داشته باشند. یادگیری عمیق بر مبنای مجموعه‌ای از الگوریتم‌ها است که ویژگی های سطح بالا در داده‌ها را با استفاده از یک شبکه عمیق که شامل چندین لایه پردازشی است، مدل‌سازی می‌کند. بدین صورت که داده‌های سطح پایین را در هر لایه دریافت می‌کند، آنها را پردازش کرده و در نتیجه داده‌های سطح بالاتر را به دست می‌آورد. مهم‌ترین هدف در این یادگیری، استخراج ویژگی‌ها توسط مدل است. بدین ترتیب، برای به دست آوردن اطلاعات مورد نیاز ماشین، لزومی به نظارت کامل انسان در هر لحظه نیست. در این حوزه، شبکه‌ها و معماری‌های مختلفی وجود دارد که از جمله این شبکه‌ها می‌توان به شبکه عصبی عمیق[6]، شبکه باور عمیق[7]، شبکه عصبی بازگشتی[8]، شبکه عصبی کانولوشنی[9] (CNN)، شبکه باقی‌مانده[10] ResNet، [11]GAN و غیره اشاره کرد.

۱-۳. شبکه موبایل نت

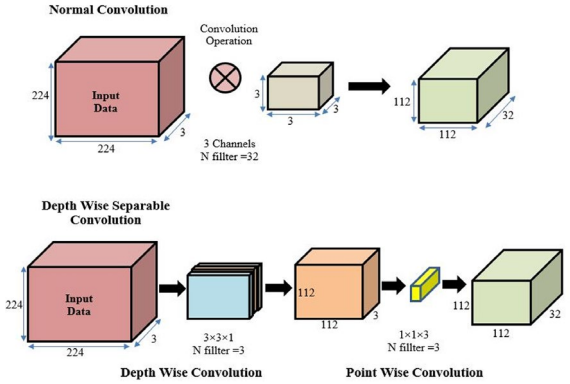
پس از پیدایش شبکه کانولوشنی AlexNet و برنده‌شدن در مسابقه ILSVRC 2012، استفاده از شبکه‌های عصبی کانولوشنی در زمینه‌های بینایی رایانه[12] افزایش یافت. با گذر زمان، شبکه‌های عمیق‌تر، پهن‌تر و البته پیچیده‌تر مانند VGG16، ResNet و غیره برای دستیابی به صحت عملکرد بالاتر مطرح شدند. با وجود افزایش پیچیدگی شبکه‌ها، سعی شده است تا تعداد پارامترهای کل شبکه کاهش یابد، اما هدف اصلی در طراحی این شبکه‌ها، افزایش دقت بوده است. به‌منظور استفاده از شبکه‌های عصبی عمیق در سیستم‌های باقدرت پردازش محدود نظیر سیستم‌های رایانه‌ای کوچک و موبایل‌ها، نیاز به تولید شبکه‌ای با اندازه کوچک و سرعت بالا همواره احساس می‌شد. بر همین اساس، دسته جدیدی از شبکه‌های کانولوشنی سبک با پارامترهای کمتر، سرعت اجرای بیشتر و البته دقت قابل‌قبول شکل گرفت. یکی از محبوب‌ترین این شبکه‌های سبک، شبکه عصبی موبایل نت نام دارد.

در شبکه‌های ساده CNN، لایهٔ کانولوشن استاندارد شامل دو مرحلهٔ فیلتر و ادغام است که طی آن در مرحلهٔ اول (فیلتر) از m عدد هسته با ابعاد k×k استفاده شده و در تصویر ورودی ضرب می‌گردد و در مرحله بعدی (ادغام)، خروجی حاصل از هر یک از m هسته[13] ، با هم جمع می‌شوند. در نتیجه، به ازای اعمال یک مرحله کانولوشن معمولی، m عدد نقشه ویژگی[14] جدید به دست می‌آید که هر یک از آن‌ها از اعمال یک هسته مجزا با ابعاد k×k به دست آمده اند. یکی از مشکلات کانولوشن استاندارد، محاسبات بالای آن است که امکان استفاده از آن را در سیستم‌های شناسایی و دسته‌بندی کاهش می‌دهد. درنتیجه، از نوع دیگری از لایهٔ کانولوشن به نام کانولوشن قابل تفکیک عمقی[15] که به محاسبات کمتری در مقایسه با کانولوشن استاندارد نیاز دارد استفاده می‌گردد [۲۶].

کانولوشن قابل‌تفکیک عمقی اساس کار شبکهٔ موبایل نت است. این نوع کانولوشن برای کاهش محاسبات از دولایه با نام‌های کانولوشن عمقی[16] و کانولوشن نقطه‌ای[17] به جای کانولوشن معمولی استفاده می‌کند. ابتدا در لایهٔ کانولوشن عمقی از یک هسته k×k استفاده شده (نتیجهٔ حاصل برخلاف کانولوشن استاندارد با هم ترکیب نمی‌شود) و سپس، لایهٔ کانولوشن نقطه‌ای از m عدد هسته c×1×1 (c تعداد کانال‌های رنگی تصویر ورودی است) به‌منظور تولید نقشه‌های ویژگی جدید استفاده می‌کند. دراین‌رابطه، k و m به ترتیب اندازه هسته کانولوشن عمقی و تعداد نقشه‌های ویژگی (کانال‌های رنگی) مربوط به خروجی می‌باشند. شکل (۱) نحوه عملکرد کانولوشن معمولی و کانولوشن

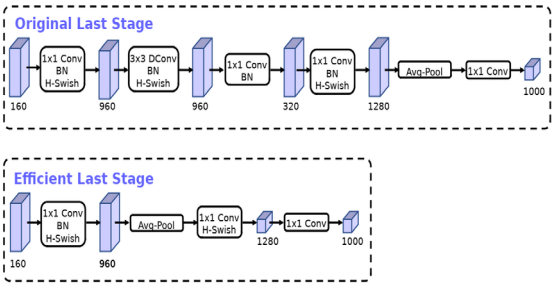
[1] Long short-term memory (LSTM)
[2] Deep Learning
[3] Machine Learning
[4] Artificial Neural Networks - ANN
[5] Deep Neural Network (DNN)
[6] Deep Neural Network
[7] Deep Belief Network
[8] Recurrent Neural Network
[9] Convolution Neural Network
[10] Residual Network
[11] Generative Adversarial Network
[12] Computer Vision
[13] Kernel
[14] Feature Map
[15] Depthwise Separable
[16] Depthwise Convolution
[17] Pointwise Convolution

قابل‌تفکیک عمقی را نشان می‌دهد. با توجه به شکل مشاهده می‌گردد که تعداد عملیات محاسباتی کاهش‌یافته است.



شکل ۱. مقایسه کانولوشن معمولی و کانولوشن تفکیک‌پذیر عمقی [۲۶].

اتصال پرشی[1] به‌منظور بهبود عملکرد شبکه در عملیات پس انتشار[2] خطا اضافه شده است. این اتصال امکان آن را فراهم می‌آورد که گرادیان از مسیرهای میان‌بر عبور کرده تا عملکرد شبکه در عملیات پس انتشار خطا بهبود یابد. به‌منظور بهبود عملکرد شبکهٔ طراحی شده، لایه های پرهزینهٔ موجود در ابتدا و انتهای شبکه بازطراحی شده و از یک تابع غیرخطی جدید، h-swish، به‌جای تابع غیر خطی ReLU استفاده شده است که تأثیر زیادی در بهبود عملکرد شبکه دارد [۲۷]. رای کاهش میزان این محاسبات و افزایش سرعت شبکه، لایهٔ ادغام میانگین[3] در بخش آخر جابه‌جا شده و قبل از لایهٔ گسترش قرار می‌گیرد. با این کار، میزان محاسبات به دلیل استفاده از لایهٔ ادغام میانگین به‌شدت کاهش می‌یابد. نتیجه طراحی به‌صورت شمای بلوکی در شکل (۲) نشان داده شده است.



شکل ۲. نمایش اصلاح لایه‌ها در مرحله آخر شبکه موبایل نت [۲۷].

یکی از توابع غیرخطی که استفاده از آن موجب بهبود دقت شبکه‌های عصبی می‌شود، توابع swish و h-swish است که از طریق معادله‌های (۱) و (۲) محاسبه می‌شوند.

$$swish[x] = x * \delta(x) \quad (۱)$$

$$h-swish[x] = x * \frac{\mathrm{Re}LU6(x+3)}{6} \quad (۲)$$

[1] Skip Connection
[2] Backpropagation
[3] Average Pooling

شبکه موبایل نت تجربه موفقی در دسته‌بندی، تشخیص و شناسایی اشیا است که توانسته با استفاده از لایه های کانولوشنی قابل تفکیک عمقی، تعداد پارامترهای شبکه کانولوشنی را کاهش داده و به دقت مناسبی نیز دست یابد. این الگوریتم توانسته بین دقت دسته‌بندی و کاهش تعداد پارامترها، تعادل خوبی برقرار نماید [۲۸].

۲. مدل پیشنهادی

روش پیشنهادی شامل چند مرحله پیش‌پردازش بر روی داده‌های سونار غیرفعال، آنالیز طیفی به‌منظور تولید طیف‌نگار مربوط به سیگنال‌های صوتی و طبقه‌بندی‌کننده موبایل نت جهت انجام‌وظیفه طبقه‌بندی است. چنین پیکربندی باعث ایجاد لایه‌هایی می‌شود که قادر به استخراج ویژگی‌های ابتدایی هستند که می‌توان آنها را در لایه‌های عمیق‌تر ترکیب کرد تا شناسایی داده‌های پیچیده را به‌خوبی انجام دهد. با استفاده از این مدل، طبقه‌بندی سیگنال سونار غیرفعال مستقیماً روی طیف‌نگارها اعمال می‌شود که از طریق تصاویر حاوی نقشه‌های طیفی قابل‌ارزیابی است.

۲-۱. مجموعه داده و پیش‌پردازش

داده‌های صوتی منتشر شده از کشتی‌ها که در این پژوهش مورد استفاده قرار گرفتند، از پایگاه داده‌ای به نام ShipsEar حاصل شده‌اند. این داده‌ها صوتی، از کشتی‌های مختلف عبوری در سواحل آتلانتیک اسپانیا ضبط شده و در پایگاه داده ShipsEar گنجانده شده است. این داده‌ها با استفاده از ضبط‌کننده‌های صوتی دیجیتال خودکار، Hyd SR-1 به دست آمده که شامل گیرنده صوت زیرآبی با حساسیت اسمی 193.5 dB در محدوده فرکانس ۱ هرتز تا ۲۸ کیلوهرتز است [۲۹].

این پایگاه‌داده از ۹۰ داده صوتی ضبط شده تشکیل شده که صداهای مربوط به ۱۱ نوع از کشتی‌ها را نشان می‌دهد. در این مقاله، زیرمجموعه‌ای از پایگاه‌داده ShipsEar تشکیل شده است که در آن ۱۱ نوع کشتی اصلی در ۴ کلاس از اندازه‌های مختلف و یک کلاس از صداهای پس‌زمینه، به‌صورت زیر ترکیب گردیدند.

کلاس A: قایق های ماهیگیری، ترال‌ها، قایق‌های صید صدف، یدک‌کش و لایروب

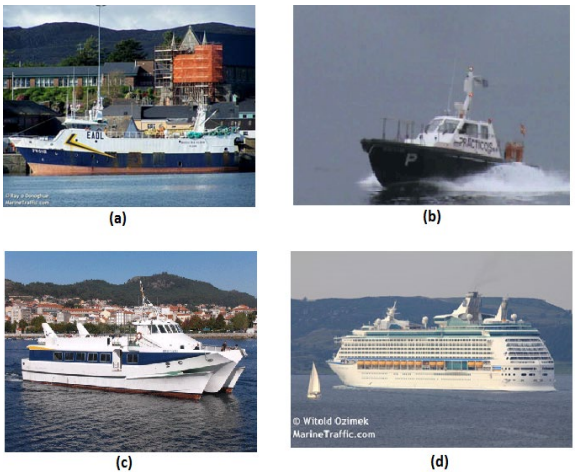
کلاس B: قایق موتوری، قایق‌های راهنما (پایلوت) و قایق‌های بادبانی

کلاس C: کشتی‌های مسافربری

کلاس D: کشتی‌های اقیانوس‌پیما

کلاس E: صداهای پس‌زمینه نویزی ضبط شده.

در شکل (۳) چهار کلاس مختلف کشتی‌ها نشان داده شده است.



شکل ۳. چهار کلاس مربوط به کشتی‌ها. (a) ترال ماهیگیری، (b) قایق پایلوت، (c) کشتی مسافربری و (d) کشتی اقیانوس‌پیما [۲۹].

اولین مرحله در پیش‌پردازش داده‌های صوتی، حذف نویزهای محیطی و انتشاری است. شناورهای سطحی از جمله منابع مولد نویز انتشاری هستند که به دلیل حرکت پروانه شناور و عمل کاویتاسیون در پشت شناور ایجاد می‌شود. اغلب این نویزها در محدوده فرکانسی ۳ کیلوهرتز می‌باشند، هرچند برخی از نویزها تا محدوده ۱۰ کیلوهرتز هم دریافت شده‌اند. در اینجا برای ازبین‌بردن این نویزها از فیلتر میانه [1] استفاده شده است.

در مرحله بعد برای آشکارکردن بخش‌های زمانی معنادار در یک سیگنال صوتی که به آن صحنه‌های شنیداری [2] نیز می‌گویند، آنها را قطعه بندی می کنند. در این فرایند ابتدا بین اجزای پراهمیت و کم اهمیت در سیگنال صوتی تفکیک پذیری انجام می‌گیرد. اجزای بااهمیت، که معمولاً آنها را با نام قطعه [3] می‌شناسیم دارای اطلاعات مهمی هستند که باید آنالیز شوند. در بین قطعات صوتی، سکوت وجود دارد که در استخراج ویژگی مورد نظر، بخش‌های سکوت حاوی اطلاعات مهمی نیستند. برای این کار از محاسبه انرژی زمان کوتاه و آستانه گذاری برای آن استفاده می گردد. تابع انرژی زمان کوتاه یک سیگنال صوتی به صورت معادله (۳) زیر تعریف می‌شود.

$$E_n = \frac{1}{N}\sum_m \left[x(m).w(n-m)\right]^2 \qquad (۳)$$

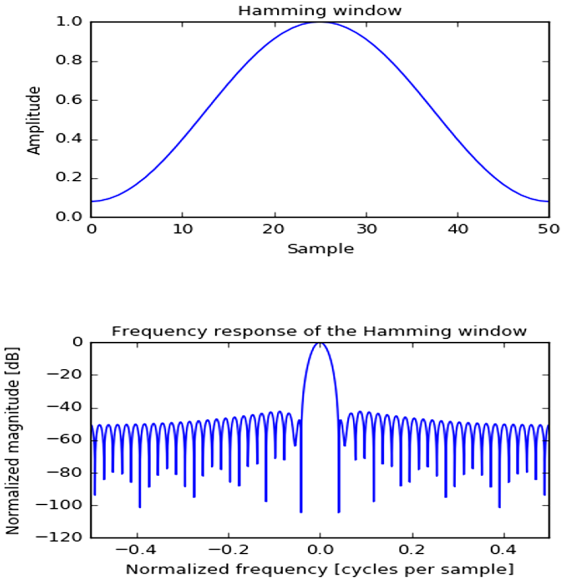
که $x(m)$ سیگنال صوتی زمان گسسته به طول N و n اندیس زمانی برای انرژی زمان کوتاه می‌باشد. w(m) معرف پنجره مستطیلی بوده که به‌صورت رابطه (۴) نشان داده می‌شود.

$$w(n) = \begin{cases} 1\,, & 0 \le n \le N-1 \\ 0, & otherwise \end{cases} \qquad (۴)$$

به‌منظور تحلیل زمان - فرکانس سیگنال با استفاده از ابزارهای در دسترس مانند تبدیل فوریه و... نیاز است که سیگنال اصلی به‌صورت سیگنال ایستا در نظر گرفته شود. به دلیل تجهیزات و ماشین‌آلات بکار رفته در کشتی‌ها سیگنال صوتی زیر آب تولید شده، یک سیگنال پویا و غیر ایستان است. بنابراین بسیاری از خواص تحلیل سیگنال نظیر تبدیل فوریه را نمی‌توان برای این دسته از سیگنال‌ها بکار برد. درواقع مشخصات آماری این سیگنال‌ها در طول زمان تغییر می‌کند و ثابت نمی‌باشد. برای حل این مشکل، با استفاده از پنجره‌سازی سیگنال، این سیگنال‌ها را به فریم‌های کوچک تبدیل می‌کنیم. این کار با این استدلال انجام می‌شود که مشخصات آماری این سیگنال‌های صوتی در بازه زمانی کوچک ثابت است و در این بازه زمانی کوچک می‌توان سیگنال را شبه ایستان در نظر گرفت و از مزایای سیگنال ایستان بهره برد. معمولاً از پنجره تحلیلی دارای همپوشانی برای درنظرگرفتن تعداد بیشتری از نمونه‌های سیگنال استفاده می‌شود. یکی از پرکاربردترین پنجره‌ها، پنجره همینگ [4] است که با معادله (۵) نشان داده شده است.

$$W(n) = 0.54 - 0.46\cos\left(2\pi/(N-1)\right) \qquad (۵)$$

همان‌طور که در شکل (۴) نشان داده شده است، پنجره همینگ یک طیف فرکانسی با باند عبور هموار و باند توقف بدون اعوجاج ایجاد می‌کند که هر دوی این خصوصیات برای به دست آوردن تخمین‌های پارامتری متغیر مهم هستند [۳۰].



شکل ۴. نمایش پنجره همینگ و پاسخ فرکانسی آن.

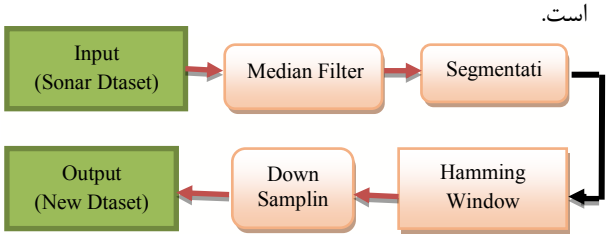
داده‌های صوتی موجود در دیتاست دارای نرخ نمونه ۵۲/۷۳۴ کیلوهرتز هستند. حال از داده‌های موردنظر به میزان ۲۶/۳۶۷ کیلوهرتز، مجدداً نمونه‌برداری می‌شود. هر چند ۵۲/۷۳۴ کیلوهرتز صدایی باکیفیت بهتر دارد، اما برای طبقه‌بندی صدا، ۲۶/۳۶۷ کیلوهرتز به‌اندازه کافی خوب است تا تفاوت صداها درک گردد. اگر اندازه هر داده نصف اندازه اصلی خود شود، مدل ایجاد

[1] Median Filter
[2] Auditory Scenes
[3] Segment
[4] Hamming Window

شده سریع‌تر خواهد بود. ۲۶/۳۶۷ کیلوهرتز دقیقاً نیمی از ۵۲/۷۳۴ کیلوهرتز است و ازآنجایی‌که میزان نمونه‌برداری به معنای تعداد دفعات نمونه‌برداری از صدای اصلی است، می‌توان از هر نمونه دیگری صرف‌نظر کرد تا نیمی از نرخ نمونه اولیه آن را به دست آورد. نحوه کاهش نرخ نمونه‌برداری در معادله (۶) نشان داده شده است.

$$x_d(n)=x_a\left(nT_s^{'}\right)=x_a\left(nMT_s^{'}\right)=x(nM) \quad (M=2) \qquad (۶)$$

در شکل (۵) نمودار بلوکی مرحله پیش‌پردازش نشان داده شده است.



شکل ۵. نمودار بلوکی پیش‌پردازش سیگنال‌های صوتی

۲-۲. استخراج طیف‌نگارها[1]

در این مرحله مجموعه داده‌های صوتی پردازش شده، با استفاده از تبدیل زمان فرکانس، مانند تبدیل فوریه، به داده‌هایی در حوزه فرکانس تبدیل می‌شوند و این امکان را فراهم می‌کند تا اطلاعات فرکانسی سیگنال صوتی ساطع شده از کشتی‌ها قابل‌استخراج باشد. به‌منظور استخراج ویژگی‌های سیگنال‌های صوتی بر اساس فرکانس، نیاز است تا با استفاده از تبدیل فوریه، سیگنال را به مقادیر فرکانس‌های آن تقسیم‌بندی نمود. زمانی که FT بر روی سیگنال صوتی اعمال می گردد، فقط مقادیر فرکانس را استخراج نموده و اطلاعات زمانی سیگنال را از دست می‌دهد. اگر از این فرکانس ها به‌عنوان ویژگی استفاده گردد، سیستم ممکن است اطلاعات ابتدایی و انتهایی را از دست دهد. روشی برای محاسب ویژگی‌های سیستم وجود دارد که مقادیر فرکانس را همراه با زمانی که در آن مشاهده شده‌اند، استخراج می‌کند. به این نمایش بصری فرکانس‌های یک سیگنال صوتی نسبت به زمان، طیف‌نگار گفته می‌شود. در نمایش طیف گرام، یک محور نشان‌دهنده زمان و محور دوم نشان‌دهنده فرکانس ها و رنگ‌ها نشان‌دهنده بزرگی (دامنه) فرکانس مشاهده‌شده در یک‌زمان خاص است [۳۱].

یکی از معایب تبدیل فوریه کلاسیک عدم انعکاس رفتارهای محلی سیگنال است. یک راه ساده برای غلبه بر این مشکل، مقایسه سیگنال با توابع پایه‌ای است که به‌صورت هماهنگ در حوزه های زمان و فرکانس محلی شده باشند. تبدیل فوریه کوتاه‌مدت[2] یا به‌اختصار STFT یکی از روش‌های اولیه مقایسه سیگنال با این توابع مقدماتی بوده و به‌صورت معادله (۷) تعریف می‌شود.

$$STFT(t,\omega)=\int_{-\infty}^{+\infty}s(\tau)\gamma_{t,\omega}^{*}(\tau)d\tau \qquad (۷)$$

در رابطه بالا، γ(t) یک تابع زمانی دارای طول محدود به نام تابع پنجره است. به‌عبارت‌دیگر، تبدیل فوریه کوتاه‌مدت، تبدیل فوریه پنجره شده سیگنال در فواصل زمانی مختلف است. این تبدیل به‌صورت معادله (۸) و (۹) نیز نشان داده می‌شود.

$$STFT(t,\omega)=\int_{-\infty}^{+\infty}s(\tau)\gamma^{*}(\tau-t)e^{-j\omega\tau}d\tau=\left\langle s(t),\gamma(\tau-t)e^{j\omega\tau}\right\rangle \qquad (۸)$$

$$\gamma_{t,\omega}(\tau)=\gamma(\tau-t)e^{j\omega\tau} \qquad (۹)$$

طیف‌نگار برای توصیف انرژی سیگنال در صفحه زمان - فرکانس، یک ماتریس دوبعدی بوده و بزرگی فرکانس را همراه با زمان برای یک سیگنال صوتی نشان خواهد داد. مجموع این طیف‌نگارها به‌عنوان تصاویری در نظر گرفته می‌شوند که سیگنال‌های صوتی سونار را به تصویر تبدیل می‌کنند. نمونه‌ای از تصاویر طیف‌نگار سیگنال صوتی سونار مربوط به ۴ کلاس از کشتی‌ها در شکل (۶) نشان داده شده است.

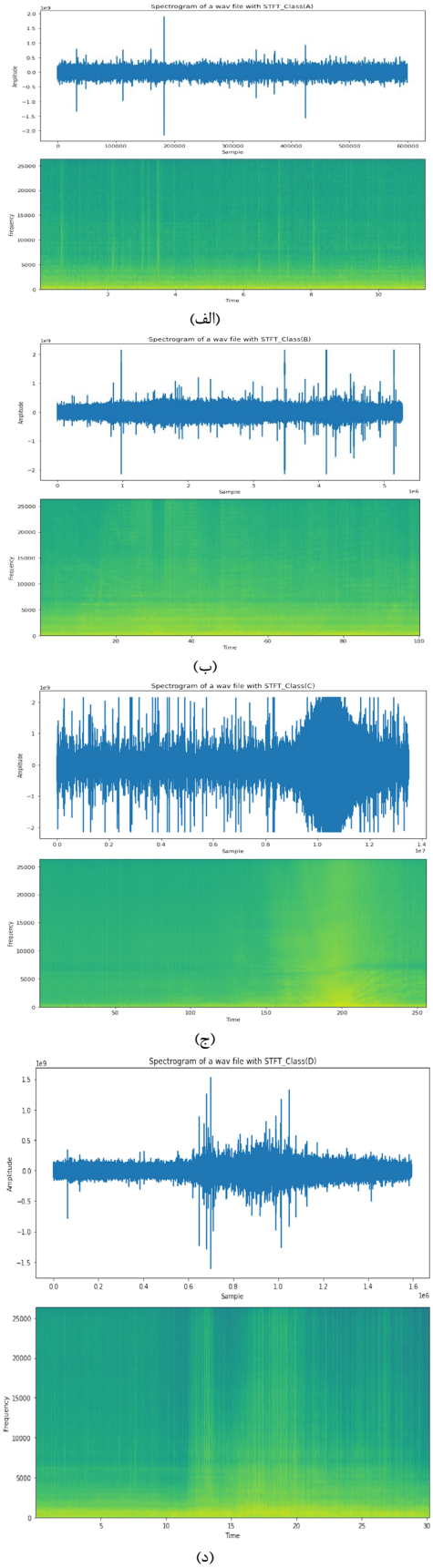
۲-۳. رویکرد فنی مدل طبقه‌بندی‌کننده

همان‌طور که در بخش قبلی گفته شد، در شبکه موبایل نت از کانولوشن قابل‌تفکیک عمقی به‌عنوان واحد اصلی استفاده می‌شود که شامل کانولوشن عمقی و کانولوشن نقطه‌ای است. ازآنجایی‌که کانولوشن عمقی یک کانولوشن تک کانالی است، تعداد نقشه‌های ویژگی[3] خروجی لایه کانولوشن عمقی در بخش میانی با نقشه‌های ویژگی ورودی یکسان است که مجموعی از نقشه‌های ویژگی خروجی تمام لایه‌های قبلی است. فیلتر کانولوشن عمقی، یک کانولوشن واحد را بر روی هر کانال ورودی انجام می‌دهد و فیلتر کانولوشن نقطه‌ای خروجی کانولوشن عمقی را به‌صورت خطی با هسته کانولوشنی[4] ۱×۱ ترکیب می‌کند. فرآیند عملکرد شبکه در شکل (۷) نشان داده شده است.

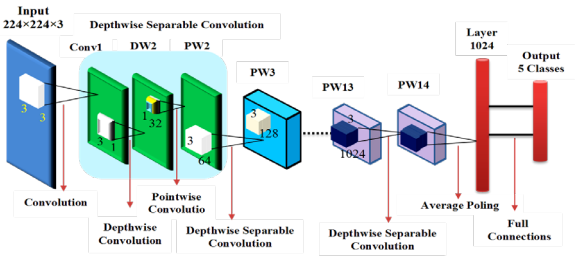
کانولوشن عمقی معادل فیلترکردن در کانولوشن استاندارد است، اما با این تفاوت که در کانولوشن استاندارد، M هسته k×k وجود داشته درحالی‌که در اینجا تنها یک هسته k×k وجود دارد. در مرحله دوم، کانولوشن نقطه‌ای قرار دارد. این مرحله معادل مرحله ادغام در کانولوشن استاندارد است. با این تفاوت که مرحله ادغام در کانولوشن استاندارد، یک جمع ساده هست، درحالی‌که مرحله ادغام در کانولوشن نقطه‌ای شامل یک کانولوشن ۱×۱ است. در اولین لایه، یک‌لایه کانولوشن استاندارد با ۳۲ فیلتر به‌اندازه ۳×۳ در نظر گرفته شده است. گام[5] کانولوشن در اولین لایه کانولوشن برابر با ۱ تنظیم شده و به‌جای استفاده از لایه ادغام بیشینه[6]، از کانولوشن با گام ۲ در پشت لایه نرمال‌سازی دسته‌ای[7] استفاده شده است.

[1] Spectrograms
[2] Short-Time Fourier Transform

[3] Feature Map
[4] Convolution Kernel
[5] Stride
[6] Maximum Pooling Layer
[7] Batch Normalization Layer

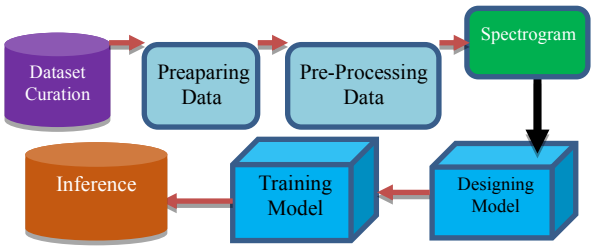


شکل ۶. تبدیل فوریه زمان کوتاه و طیف‌نگار مربوط به سیگنال‌های صوتی سونارى. الف) کلاس A ، ب) کلاس B، ج) کلاس C و د) کلاس D.



شکل ۷. فرآیند شبکه موبایل نت در روش پیشنهادی.

از لایه دوم به بعد تمام لایه‌ها از نوع کانولوشن DWS هستند. لایه کانولوشن عمقی با علامت DW و نقطه‌ای با علامت PW و سایز ۱×۱ نشان داده شده است. اندازه فیلتر تمامی لایه‌های کانولوشن عمقی ۳×۳ هست. سه‌لایه کانولوشن با ۳۲ فیلتر وجود دارند و پس از آن یک‌لایه کانولوشن نقطه‌ای با ۶۴ فیلتر در نظر گرفته شده است. سپس ۹ لایه با ۱۲۸ فیلتر قرار گرفته است. در پایان هم ۲ لایه ۱۰۲۴ فیلتری وجود خواهد داشت. ساختار کلی مدل پیشنهادی در شکل (۸) نشان داده شده است.



شکل ۸. نمای کلی بخش‌های مختلف مدل پیشنهادی.

۳. بخش تجربی

در این تحقیق ویژگی‌های استخراج شده از داده‌های سونارى در حوزه زمان - فرکانس در قالب تصاویر طیف‌نگار ذخیره می‌شوند. این تصاویر به‌عنوان ورودی مدل طبقه‌بندی‌کننده پیشنهادی مورد استفاده قرار می‌گیرند. به‌منظور اجرای آزمایش‌های در نظر گرفته شده، داده‌ها با فرکانس نمونه‌برداری ۲۶/۳۶۷ کیلوهرتز دوباره نمونه‌برداری می‌شوند. هر یک از فایل‌های صوتی به بخش‌های[۱] متعددی تقسیم می‌شوند تا برای ورودی الگوریتم‌های یادگیری عمیق پردازش شوند. طول خاص بخش بندی بر اساس ماهیت سیگنال به دست آمده و ابعاد ورودی الگوریتم مورد استفاده، تعیین می‌شود. با درنظرگرفتن ویژگی‌های سیگنال‌های صوتی سونار غیرفعال، منابع محاسباتی و دقت طبقه‌بندی، هر سیگنال به بخش‌های ۴ ثانیه‌ای تقسیم کردیم. هر بخش توسط مدل به‌صورت مستقل طبقه‌بندی می‌شود

با انجام عمل بخش‌بندی و محاسبه طیف فرکانسی، ۵۶۷۱ تصویر طیف‌نگار در ابعاد ۳×۲۲۴×۲۲۴ به دست می‌آید که متعلق به ۵ کلاس تعریف شده از انواع شناورها است. ۷۰٪ از این داده‌ها

[۱] Segment

برای آموزش[1]، ۲۰٪ برای اعتبارسنجی[2] و ۱۰٪ برای آزمایش[3] استفاده شده است. عملکرد الگوریتم‌ها با پارامترهایی مانند صحت[4]، دقت[5]، بازیابی[6] و امتیاز F1[7] ارزیابی می‌شود. معادلات ارزیابی از طریق روابط (۱۰) تا (۱۵) قابل‌محاسبه می‌باشند [۳۲].

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \quad (۱۰)$$

$$False\ Positive\ Rate = \frac{FP}{TN + FP} \quad (۱۱)$$

$$Accuracy = \frac{True\ Positives + True\ Negatives}{TP + TN + FP + FN} \quad (۱۲)$$

$$precision = \frac{True\ Positives}{True\ Positives + False\ Positive} \quad (۱۳)$$

$$\mathrm{Re}call = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (۱۴)$$

$$F1 - Score = 2 \times \frac{Prcision \times Recall}{Prcision + Recall} \quad (۱۵)$$

با استفاده از ماتریس پریشانی[8] میزان اطمینان و صحت طبقه‌بندی در کلاس‌های مختلف نشان داده می‌شود. هدف از انجام آزمایش‌ها مختلف، یافتن بالاترین نرخ مثبت صحیح[9] و کمترین نرخ مثبت کاذب[10]، خواهد بود. در جدول (۱)، نحوه نمایش اطلاعات در ماتریس پریشانی نشان داده شده است.

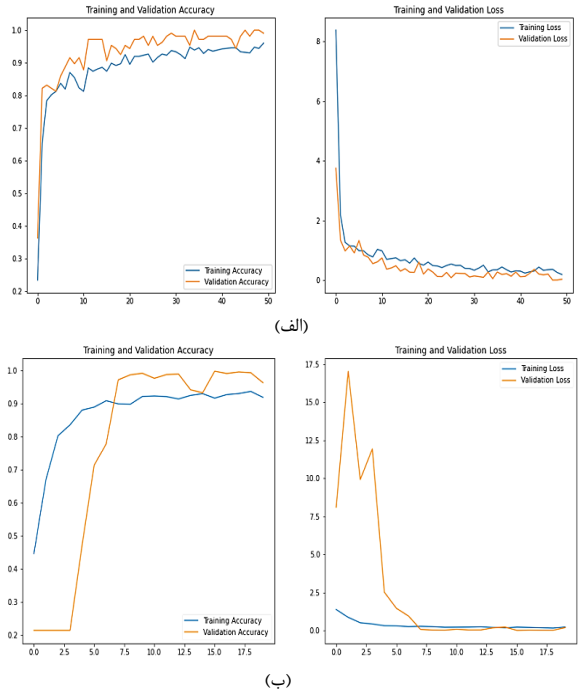
جدول ۱. اطلاعات موجود در ماتریس پریشانی

		Actual Value	
		True	False
Predicted Value	True	True Positive (Correect Detection)	False Negative (False Detection)
	False	False Positive (False Alarm)	True Negative (Correct Detection)

مدل طراحی شده بر روی داده‌های آموزشی، با دسته‌بندی ۶۴تایی[11] و تعداد ۲۰ و ۵۰ تکرار[12] آموزش داده می‌شود. Batch بیانگر تعداد نمونه‌های آموزشی موجود در یک دسته واحد و Epochs نشان‌دهنده تعداد تکرارها در حالتی است که یک مجموعه داده کامل از طریق شبکه عصبی یک‌بار به جلو و یک‌بار به عقب منتقل می‌شود. در این تحقیق، چهار مدل مبتنی بر یادگیری عمیق (شبکه کانولوشنی CNN استاندارد [۱۴]، شبکه VGG [۱۱]، شبکه ResNet18 [۲۳] و شبکه LSTM [۲۵]) بر اساس محبوبیت و استفاده مکرر آنها در مسئله طبقه‌بندی آکوستیک زیر آب انتخاب گردیده و نتایج روش پیشنهادی با آنها مقایسه شده است.

۴. نتایج و بحث

این مدل در پایتون با استفاده از سری کتابخانه Keras با پشت پرده[13] Tensorflow ساخته شده است. اندازه دوره‌ها برابر ۲۰ و ۵۰ و همچنین اندازه دسته برابر ۶۴ انتخاب شده است. صحت و زیان مدل را هم برای داده‌های آموزشی و هم برای داده‌های اعتبارسنجی به دست می‌آوریم. نمودار تابع هزینه و دقت مدل در هر دوره در شکل (۹) نشان داده شده است.

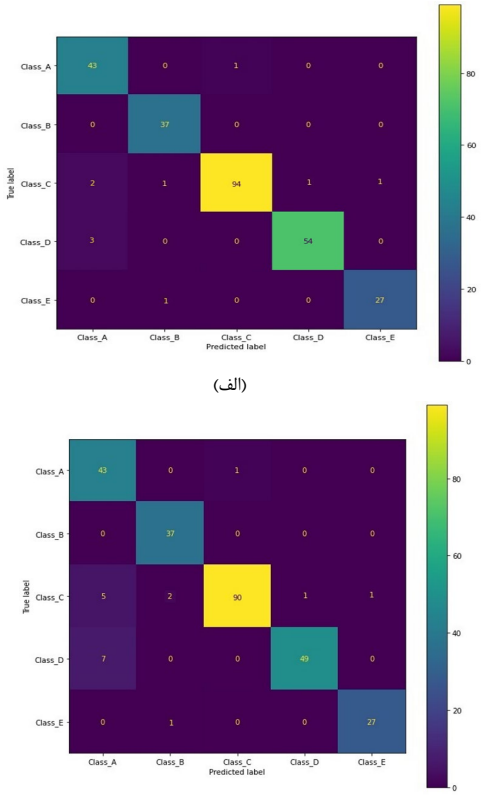


شکل ۹. نمایش صحت آموزش و زیان ارزیابی در روش پیشنهادی الف) اندازه دوره برابر ۵۰ و ب) اندازه دوره برابر ۲۰.

ماتریس پریشانی برای نمونه داده‌های آزمایش در شکل (۱۰) نشان داده شده است. قطر ماتریس نشان‌دهنده نتایج حاصل از بازیابی یا نرخ مثبت واقعی است که عملکرد صحیح مدل را بر اساس طبقه‌بندی صحیح کلاس‌های مختلف مجموعه داده بیان می‌کند.

یکی از روش‌های ارزیابی توانایی مدل در شناسایی کلاس‌های مختلف اهداف در یادگیری عمیق، استفاده از منحنی ROC است. این منحنی به وسیله ترسیم نرخ مثبت صحیح (TPR) نسبت به نرخ مثبت کاذب (FPR) ایجاد می‌شود. نسبت شناسایی صحیح داده‌های آزمایشی مربوط به کلاس‌های مختلف توسط منحنی ROC در شکل (۱۱) نشان داده شده است.

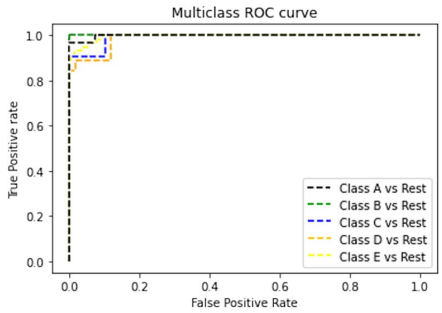
[1] Train
[2] Validation
[3] Test
[4] Accuracy
[5] Precision
[6] Recall
[7] F1-Score
[8] Confusion Matrix
[9] True Positive Rate
[10] False Positive Rate
[11] Batch Size=64
[12] Epochs=20, 50
[13] Backend



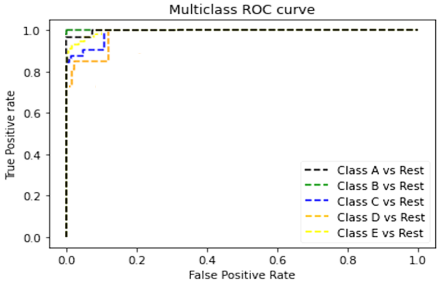
(الف)

(ب)

شکل ۱۰. نمایش ماتریس پریشانی مجموعه داده آزمایشی الف) اندازه دوره برابر ۵۰ و ب) اندازه دوره برابر ۲۰.



(الف)



(ب)

شکل ۱۱. مقایسه شناسایی صحیح کلاس‌های مختلف با منحنی ROC ، الف) اندازه دوره برابر ۵۰ و ب) اندازه دوره برابر ۲۰.

نتایج دقت شناسایی سیگنال‌های آکوستیکی برای همه روش‌ها در جدول (۲) آورده شده است. نتایج نشان می‌دهد شبکه پیشنهادی به مقدار درستی شناسایی ۹۶/۲۷ درصد دست‌یافته که از روش‌های CNN استاندارد، VGG19، ResNet و LSTM بهتر عمل کرده است. همچنین دقت ۹۸/۳۷ درصد، بازیابی ۹۹ درصد و امتیاز F1 ۹۹ درصد از دیگر نتایج حاصله است. با بررسی نتایج مشاهده می‌گردد روش پیشنهادی در تمام معیارهای ارزیابی برای شناسایی دقیق و صحیح اهداف شناور سطحی بر اساس سیگنال‌های صوتی دریافت شده از آنان، نسبت به سایر روش‌های متداول در شناسایی خودکار اهداف، دارای عملکرد مناسب‌تر و قابل‌اعتمادی است.

جدول ۲. مقایسه نتایج حاصل از الگوریتم‌های مختلف یادگیری عمیق

Method	Accuracy	Precision	Recall	F1-Score
CNN	٪۸۷/۲۰	٪۹۴/۷۵	٪۹۵/۰۸	٪۸۷/۶۵
VGG 19	٪۷۸/۱۳	٪۸۱/۰۰	٪۷۶/۱۰	٪۸۳/۴۰
LSTM	٪۹۴/۹	٪۹۵/۸	٪۹۳/۷	٪۹۴/۸
ResNet18	٪۹۴/۶۰	٪۹۴/۳۰	٪۹۴/۲۰	٪۹۳/۸۰
Proposed	٪۹۶/۲۷	٪۹۸/۳۷	٪۹۹/۰۰	٪۹۹/۰۰

جدول (۳) بازده محاسباتی مدل پیشنهادی را با توجه به تعداد عملیات و زمان محاسبات انجام شده در هر دوره و کل دوره، را نشان می‌دهد. نتایج با استفاده از سیستمی با پردازنده i7-7700HQ @2.8 گیگاهرتز، NVIDIA GeForce GTX 1050 Ti به دست آمده است. مشاهده می‌شود که تعداد پارامترهای عملیاتی روش پیشنهادی به‌مراتب کمتر از مدل‌های مبتنی بر CNN، VGG19 و ResNet است. این مهم به دلیل استفاده از روش ادغام میانگین و حذف کانولوشن‌های اضافی نقطه‌ای و عمقی اضافی در انتهای الگوریتم پیشنهادی است. مدت‌زمان انجام محاسبات آموزش و ارزیابی، صرف شده توسط روش پیشنهادی، کمتر از مدل‌های طبقه‌بندی‌کننده مبتنی بر روش‌های کانولوشنی مورد اشاره است.

جدول ۳. مقایسه تعداد پارامترها و مدت‌زمان انجام محاسبات در الگوریتم‌های مختلف یادگیری عمیق

Method	Num of Parameters	Computation Time (sec/ epoch)	Computation Time (Total)
CNN	۴۷ میلیون	۷۰ ثانیه	۱۰۳ دقیقه
VGG 19	۲۲/۵ میلیون	۷۷ ثانیه	۱۱۸/۵ دقیقه
ResNet	۲۳/۸ میلیون	۷۹ ثانیه	۱۲۲/۵ دقیقه
Proposed	۵/۸ میلیون	۶۸ ثانیه	۹۰/۶ دقیقه

با توجه به نقش دوره انجام آموزش در کاهش خطای ارزیابی و افزایش دقت شناسایی مدل، در این تحقیق از دوره‌های زمانی مختلف (Epoch=20,50) برای آموزش داده استفاده شده که نتایج حاصل در جدول (۴) آورده شده است. در واقع به زمانی که تمام داده‌های آموزشی به طور هم‌زمان مورد استفاده قرار گیرند یا به‌عنوان تعداد کل تکرار تمام داده‌های آموزشی در یک‌چرخه برای آموزش مدل یادگیری عمیق، یک دوره گفته می‌شود. هر نمونه در

مجموعه داده آموزشی یک‌بار در طول یک دوره فرصت به‌روزرسانی پارامترهای مدل داخلی را خواهد داشت.

جدول ۴. نتایج حاصل از آموزش مدل در Epoch های مختلف

Method	Num of Epochs	Accuracy	Loss
CNN	۲۰	٪۷۹/۴۵	٪۱۱/۵
	۵۰	٪۸۷/۲۰	٪۱۰/۴
VGG 19	۲۰	٪۹۰/۸۱	٪۹/۶
	۵۰	٪۹۳/۰۲	٪۸/۲
ResNet	۲۰	٪۸۸/۱۰	٪۱۲/۶
	۵۰	٪۸۸/۱۲	٪۱۰/۵
Proposed	۲۰	٪۹۶/۱۳	٪۳/۲
	۵۰	٪۹۶/۲۷	٪۲/۰۱

۵. نتیجه‌گیری

در این مقاله، از یک مجموعه داده آکوستیکی زیر آب که شامل صداهای صوتی منتشر شده از کلاس‌های مختلف کشتی‌ها است، استفاده شده است. در این تحقیق یک مدل شبکه عصبی کانولوشنی عمیق مبتنی بر روش یادگیری عمیق (موبایل نت) با مکانیسم‌های تغییر داده شده در ابتدا و انتهای شبکه برای شبیه‌سازی پردازش داده‌های صوتی سونار و تبدیل آنها به تصاویر طیف‌نگار به‌منظور شناسایی و طبقه‌بندی انواع کشتی‌ها پیشنهاد شده است. در راستای سرعت‌بخشیدن به مراحل شناسایی و طبقه‌بندی اهداف توسط مدل، پیش‌پردازش‌های صوتی بر روی داده‌های آکوستیکی انجام شده و در نهایت از شبکه کانولوشنی بازطراحی شده موبایل نت، با کمترین تعداد پارامتر استفاده شده است. نتایج آزمایش‌ها طبقه‌بندی نشان می‌دهد که دقت روش پیشنهادی نسبت به روش‌های متداول یادگیری عمیق در طبقه‌بندی تصاویر، بهبود نسبی داشته و از نظر سرعت محاسبات و زیان ارزیابی، بهتر عمل کرده است. این شبکه می‌تواند با تولید طیف‌نگارهای مطابق با داده‌های آکوستیکی که با ماهیت سیگنال‌های منتشر شده از کشتی‌ها سازگار هستند، ویژگی‌های مبتنی بر زمان و فرکانس داده‌ها را به طور هم‌زمان استخراج نماید. با استفاده از روش پیشنهادی می‌توان با ایجاد شبکه‌ای از گیرنده‌های صوتی زیر آب، ضمن جمع‌آوری سیگنال‌های صوتی زیر آب، از طریق پردازش سیگنال‌های آکوستیکی زیر آب، به‌صورت هوشمند و بدون نیاز به کاربر انسانی، شناورهای سطحی عبوری را شناسایی و هشدارهای لازم را ارسال نمود. از این شبکه نه‌تنها برای خدمات نظامی، بلکه می‌توان در امداد و نجات دریایی، سیستم‌های کنترل و نظارت زیر آب برای جلوگیری از نفوذهای زیر آب به‌وسیله غواصان و شناورهای بدون سرنشین زیر آبی و همچنین صنعت شیلات و ماهیگیری بهره‌برداری شود. در این مقاله سعی شده است با فشرده‌سازی و کاهش میزان محاسبات (با طراحی و یا اصلاح لایه‌های موجود) و عمیق‌تر کردن این نوع از شبکه های کانولوشنی، به توازن مناسبی بین دقت و سرعت شبکه‌ها دست یابند. در مقایسه با مدل‌های متداول یادگیری عمیق، ضمن ارتقای نسبی دقت طبقه‌بندی، تعداد پارامترها به طور محسوسی کاهش‌یافته و مقدار محاسبات را کم شده است. به‌طورکلی، مدل‌های ارائه شده در این مقاله قابلیت اجرا بر روی دستگاه‌های قابل‌حمل و دستگاه تلفن همراه خواهد داشت.

۶. مرجع‌ها

[1] Hu, G.; Wang, K.; Peng, Y.; Qiu, M.; Shi J.; Liu, L. “Underwater Acoustic Target Recognition Based on Supervised Feature-Separation Algorithm”; Comput. Intell. Neurosci. 2018, Article ID 1214301, 10 pages.

[2] Yan, J.; Sun, H.; Chen, H.; Junejo, N. U. R.; Cheng, E. “Resonance-Based Time-Frequency Manifold for Feature Extraction of Ship-Radiated Noise”; Sensors 2018, 18, 936.

[3] Neupane, D.; Seok, J. “A Review on Deep Learning-Based Approaches for Automatic Sonar Target Recognition”; J. Electronics 2020, 9, 1972.

[4] Howard, A.; Sandler, M.; Chu, G.; Chen, L.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V. “Searching for Mobilenetv3”; Proc. IEEE Int. Conf. Comput. Vis. CVF, 2019, 1314-1324

[5] Gorman, R. P.; Sejnowski, T. J. “Learned Classification Of Sonar Targets Using a Massively Parallel Network”; IEEE Trans. Acoust. 1988, 36, 1135-1140. https://doi.org/10.1109/29.1640.

[6] Chin-Hsing, C.; Jiann-Der, L.; Ming-Chi, L. “Classification of Underwater Signals Using Wavelet Transforms and Neural Networks”; Math. Comput. Model, 1998, 27, 47-60. https://doi.org/10.1016/S0895-7177(97)00259-8.

[7] Azimi-Sadjadi, M. R.; Yao, D.; Dobeck, G. “Underwater Target Classification Using Wavelet Packets and Neural Networks”; IEEE Trans. Neural. Netw. 2000, 11, 784-794.

[8] Baran, R. H.; Coughlin, J. P. “A Neural Network for Target Classification Using Passive Sonar”; IEEE ACM. 1991, 188–198.

[9] Williams, D. P. “Underwater Target Classification in Synthetic Aperture Sonar Imagery Using Deep Convolutional Neural Networks”; 23rd IEEE. ICPR, 2016, 2497–2502.

[10] Galusha, A.; Dale, J.; Keller, J.; Zare, A.; “Deep Convolutional Neural Network Target Classification for Underwater Synthetic Aperture Sonar Imagery”; Proc. of SPIE. 2019,11012. 1101205.

[11] Bach, N. H.; Vu, L. H.; Nguyen, V. D. “Classification of Surface Vehicle Propeller Cavitation Noise Using Spectrogram Processing in Combination with Convolution Neural Network”; Sensors 2021, 21, 1-18.

[12] Yang, H.; Shen, S.; Yao, X.; Sheng, M.; Wang, C. “Competitive Deep-Belief Networks for Underwater Acoustic Target Recognition”; Sensors 2018, 18, 952.

[13] Park, J.; Jung, D. J. “Identifying Tonal Frequencies in a Lofargram with Convolutional Neural Networks”; ICCAS. 2019, 19, 338-341. https://doi.org/10.23919/ICCAS47443.2019.8971701.

[14] Yang, H.; Xue, L.; Hong, X.; Zeng, X. “A Lightweight Network Model Based on an Attention Mechanism for Ship-Radiated Noise Classification”; J. Mar. Sci. Eng. 2023, 11, 1-17. https://doi.org/10.3390/jmse11020432

[15] Miao, Y.; Zakharov, Y.; Sun, H.; Li, J.; Wang, J. "Underwater Acoustic Signal Classification Based on Sparse Time–Frequency Representation and Deep Learning"; IEEE J. Ocean. Eng. 2021, 46, 952-962. https://doi.org/10.1109/JOE.2020.3039037

[16] Yang, H.; Sheng, S.; Yao, X.; Li, J.; Xu, X.; Sheng, M. "Ship Type Classification by Convolutional Neural Networks with Auditory-Like Mechanisms"; Sensors 2020, 20, 253.

[17] Jin, A.; Zeng, X. "A Novel Deep Learning Method for Underwater Target Recognition Based on Res-Dense Convolutional Neural Network with Attention Mechanism"; J. Mar. Sci. Eng. 2023, 11, 1-20.

[18] Jiang, Z.; Zhao, C.; Wang, H. "Classification of Underwater Target Based on S-ResNet and Modified DCGAN Models"; Sensors 2022, 22, 2293.

[19] Tian, S.; Chen, D.; Wang, H.; Liu, J. "Deep Convolution Stack for Waveform in Underwater Acoustic Target Recognition"; Sci. Rep. 2021, 11, 9614.

[20] Hu, G.; Wang, K.; Liu, L. "Underwater Acoustic Target Recognition Based on Depthwise Separable Convolution Neural Networks"; Sensors 2021, 21, 1429.

[21] Chen, J.; Chang, L.; Zhang, J.; Han, B. "Underwater Target Recognition based on Multi-Decision LOFAR Spectrum Enhancement: A Deep Learning Approach"; Future Internet 2021, 13, 265.

[22] Saffari, A.; Zahiri, S. H.; Khozein Ghanad, N. "Using SVM Classifier and Micro-Doppler Signature for Automatic Recognition of Sonar Targets"; Archives of Acoustics 2023, 48, 49–61.

[23] Hong, F.; Liu, C.; Guo, L. "Underwater Acoustic Target Recognition with ResNet18 on ShipsEar Dataset"; 4th International Conference on Electronics Technology, Chengdu, China, 2021, 1240–1244.

[24] Li, L.; Song, S.; Feng, X. "Combined LOFAR and DEMON Spectrums for Simultaneous Underwater Acoustic Object Counting and F0 Estimation"; J. Mar. Sci. Eng. 2022, 10, 1565.

[25] Kamal, S.; Chandran, C. S.; Supriya, M. H. "Passive Sonar Automated Target Classifier for Shallow Waters Using End-to-End Learnable Deep Convolutional LSTMs"; Engineering Science and Technology 2021, 1-12.

[26] Ke, X.; Yuan, F.; Cheng, E. "Deep Learning Methods for Underwater Target Feature Extraction and Recognition"; Sensors 2018, 18, 4318.

[27] Kulkarni, U.; Meena, S. M.; Sunil, V.; Gopal, B. "Quantization Friendly MobileNet (QF-MobileNet) Architecture for Vision Based Applications on Embedded Platforms"; Neural. Networks. 2021, 136, 28–39.

[28] Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications"; Cite Arxiv:1704.04861. CVPR, 2017.

[29] Santos-Domínguez, D.; Torres-Guijarro, S.; Cardenal-López, A.; Pena-Gimenez, A. "ShipsEar: an Underwater Vessel Noise Database"; Appl Acoust. 2016, 113, 64-69.

[30] Tan, L.; Jiang, J. "Digital Signal Processing"; 3rd edition. England, 2019, Chapter7, 229-313.

[31] Neupane, D.; Seok, J. "A Review on Deep Learning-Based Approaches for Automatic Sonar Target Recognition"; J. Electronics 2020, 9, 1972.

[32] Irfan, M.; Jiangbin, Z.; Ali, S.; Iqbal, M.; Masood, Z.; Hamid, U. "DeepShip: An Underwater Acoustic Benchmark Dataset and a Separable Convolution Based Autoencoder for Classification"; Expert. Syst. Appl. 2021, 183, 115270. https://doi.org/10.1016/j.eswa.2021.115270