

Sistema FIEB



CENTRO UNIVERSITÁRIO SENAI CIMATEC

PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM

COMPUTACIONAL E TECNOLOGIA INDUSTRIAL

Mestrado em Modelagem Computacional e Tecnologia Industrial

Defesa de Mestrado

**Modelo computacional para reconhecimento de objetos
em ambiente subaquático baseado em aprendizagem de
máquina**

Apresentada por: Igor Vilas-Bôas Silveira
Orientador: Dr. Roberto Luiz Souza Monteiro
Co-orientador: Dr. Oberdan Rocha Pinheiro

Novembro de 2023

Igor Vilas-Bôas Silveira

**Modelo computacional para reconhecimento de objetos
em ambiente subaquático baseado em aprendizagem de
máquina**

Defesa de Mestrado apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Curso de Mestrado em Modelagem Computacional e Tecnologia Industrial do CENTRO UNIVERSITÁRIO SENAI CIMATEC, como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Área de conhecimento: Modelagem de Processos Industriais

Orientador: Dr. Roberto Luiz Souza Monteiro

CENTRO UNIVERSITÁRIO SENAI CIMATEC

Co-orientador: Dr. Oberdan Rocha Pinheiro

CENTRO UNIVERSITÁRIO SENAI CIMATEC

Ficha catalográfica elaborada pela Biblioteca do Centro Universitário SENAI CIMATEC

S587m Silveira, Igor Vilas-Bôas

Modelo computacional para reconhecimento de objetos em ambiente subaquático baseado em aprendizagem de máquina / Igor Vilas-Bôas Silveira. – Salvador, 2024.

63 f. : il. color.

Orientador: Prof. Dr. Roberto Luiz Souza Monteiro.

Coorientador: Prof. Dr. Oberdan Rocha Pinheiro.

Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) – Programa de Pós-Graduação, Centro Universitário SENAI CIMATEC, Salvador, 2024. Inclui referências.

1. Fundo do mar. 2. Exploração - Mar. 3. Reconhecimento de objetos. 4. Aprendizagem industrial. I. Centro Universitário SENAI CIMATEC. II. Monteiro, Roberto Luiz Souza. III. Pinheiro, Oberdan Rocha. IV. Título.

CDD 620.00113

CENTRO UNIVERSITÁRIO SENAI CIMATEC**Mestrado Acadêmico em Modelagem Computacional e Tecnologia Industrial**

A Banca Examinadora, constituída pelos professores abaixo listados, aprova a Defesa de Mestrado, intitulada “**Modelo computacional para reconhecimento de objetos em ambiente subaquático baseado em aprendizagem de máquina**” apresentada no dia 21 de novembro de 2023, como parte dos requisitos necessários para a obtenção do Título de Mestre em Modelagem Computacional e Tecnologia Industrial.

Assinado eletronicamente por:
Roberto Luiz Souza Monteiro
CPF: ***.881.935-**
Data: 28/12/2023 17:09:00 -03:00

Orientador:

Prof. Dr. Roberto Luiz Souza Monteiro
SENAI CIMATEC

Assinado eletronicamente por:
Oberdan Rocha Pinheiro
CPF: ***.073.705-**
Data: 28/11/2023 09:10:37 -03:00

Coorientador:

Prof. Dr. Oberdan Rocha Pinheiro
SENAI CIMATEC

Electronically signed by:
Taniel Silva Franklin
CPF: ***.769.875-**
Date: 11/27/2023 7:39:49 PM -03:00

Membro Externo:

Prof. Dr. Taniel Silva Franklin
FIEB

Assinado eletronicamente por:
Guilherme de Melo Baptista Domingues
CPF: ***.888.927-**
Data: 27/11/2023 20:03:39 -03:00

Membro Externo:

Prof. Dr. Guilherme de Melo Baptista Domingues
UERJ

Dedico este trabalho a Loíze Raquel (*in memorian*), a pessoa mais incrível que cruzou o meu caminho nessa minha jornada aqui neste planeta. Querida Loíze, a força da sua presença ainda persiste a reverberar em todos os lugares. Seu amor pulsa em meu coração e arde como fogo sagrado. Sei que você me aguarda em outra dimensão e que nossa conexão é de alma. Mas confesso que não tem sido nada fácil seguir adiante. Dedico este trabalho a ti, pois sei o quanto você lutou para que eu não desistisse não importassem os percalços e tormentas na estrada. Obrigado por tudo Lolo.

Agradecimentos

Primeiramente, deixo do meus sinceros votos de agradecimento aos professores componentes da banca: professor Dr. Roberto Monteiro por ter me recebido como orientando e me apoiado na importante etapa de qualificação, ao professor Dr. Oberdan pelas grandes contribuições nas simulações e ensinamentos durante toda minha jornada de escrita da dissertação, professor Dr. Guilherme e professor Dr. Taniel pelos apontamentos e contribuições para melhoria deste trabalho.

Agradeço também o meu amigo Osmar pelos conselhos e direcionamentos no momento em que tive dificuldade com a dissertação e ao professor Dr. Danilo (UFRJ) que me ofereceu a oportunidade única de realizar o meu estágio docência na UFRJ.

Não posso deixar de lembrar, também, do meu amigo Gustavo Licursi, pelos debates sobre neurociência que ampliaram a minha visão sobre as redes neurais naturais, bem como pelas suas contribuições na seção [2.2](#) e na produção das imagens das figuras [2.3](#) e [2.2](#).

Por fim, deixo meus votos de gratidão à pessoa aqui homenageada, por ter sido a minha maior incentivadora nessa jornada e verdadeiro motivo por eu ter decidido reotmar o rumo deste trabalho e ir até o final como uma forma de homenageá-la e demonstrar que guardo com carinho os ensinamentos que dela recebi.

Salvador, Brasil

Igor Vilas-Bôas Silveira

25 de Novembro de 2023

Resumo

Desde a antiguidade, o homem tem explorado a superfície das águas do planeta realizando trocas comerciais, intercambiando valores e costumes, e permeando prosperidade. Ao expandir suas fronteiras para o além-mar, o homem experimentou uma interconexão global, onde uma malha invisível de inter-relações complexas conecta pessoas de todo mundo. Apesar da intimidade criada com as navegações dos oceanos, pouco se sabe ainda sobre o fundo do mar. Dificuldades impostas pela mãe natureza para se adentrar em águas profundas aliadas às limitações do homem têm corroborado para manter o ambiente subaquático pouco explorado. Entretanto, um momento ímpar para a exploração do fundo do mar se inicia, um vez que diversas tecnologias em pleno desenvolvimento convergem para a superação dos problemas de acesso ao ambiente subaquático, assim como da captura e processamento de informação desse ambiente. A interpretação das informações capturadas abaixo d'água engloba a questão da detecção e reconhecimento de objetos que é a questão central deste trabalho. A aprendizagem de máquina, nos últimos anos, tem trazido uma nova perspectiva nos métodos de reconhecimento de objetos. Recentemente, diversos trabalhos de aprendizagem de máquina e aprendizagem profunda têm sido publicados com resultados muito relevantes. Para este trabalho, uma rede de perceptrons multi camadas (MLP) foi desenvolvida para detecção de objetos abaixo d'água a partir de uma combinação de três datasets pré-existentes, resultando em um conjunto de dados com três classes distintas que englobam objetos naturais (tubarões e tratarugas) e objetos artificiais (tubos de tamanhos diversos). As simulações da MLP apresentaram como resultados um número de 888 instâncias classificadas corretamente, contra apenas 12 incorretas, alcançando uma acurácia de 98,7%. Os resultados alcançados nesse trabalho de investigação evidenciam que o modelo computacional proposto é viável para reconhecimento de objetos em ambiente subaquático baseado em aprendizagem de máquina.

Palavras chave: fundo do mar, exploração, reconhecimento de objetos, aprendizagem de máquina

Sumário

1	Introdução	1
1.1	Definição do Problema	1
1.2	Objetivos	2
1.3	Justificativa	2
1.4	Organização da dissertação	3
2	Revisão da Literatura	4
2.1	Representação e reconhecimento de objetos: um modelo	4
2.2	Visão humana: um exemplo de modelo de reconhecimento de objeto	5
2.3	Sistema óptico artificial, imagem digital e visão computacional	6
2.4	Detecção, representação e reconhecimento no ambiente subaquático	9
2.4.1	O veículo	10
2.4.2	O sensor e o ambiente subaquático	12
2.4.3	Processamento do sinal, interpretação e reconhecimento	13
2.5	Reconhecimento e aprendizagem de máquina	16
2.6	As redes neurais e o perceptron	17
2.7	A rede de perceptrons Multi camada	25
2.8	A rede convolucional	33
2.9	Métodos de detecção de múltiplos objetos	36
2.10	DL, ATR e a questão dos datasets	38
2.11	Por que devemos olhar para o fundo do mar?	39
2.11.1	Cabos submarinos: o mundo conectado	40
2.11.2	Atividades econômicas	41
2.11.3	Arqueologia submarina	42
2.11.4	Busca e Salvamento no mar	44
2.11.5	Detecção de minas submarinas	45
3	Materiais e métodos	48
3.1	<i>Dataset</i>	48
3.2	Modelo computacional de detecção e classificação	49
3.3	Método de validação do modelo	51
4	Resultados e discussão	53
5	Considerações Finais	58
5.1	Conclusões	58
5.2	Contribuições	58
5.3	Atividades Futuras de Pesquisa	58
	Referências	59

Lista de Tabelas

2.1	Alguns dos principais parâmetros de um RADAR	15
2.2	Tabela verdade da função XOR	26
2.3	Principais hiperparâmetros de uma MLP	32
4.1	Matriz de confusão da classificação dos dados.	56

Lista de Figuras

2.1	Diagrama simplificado: processo de representação do ambiente	4
2.2	Visão Humana: Captura de informação do ambiente	5
2.3	Visão humana: fluxo ventral	5
2.4	Olho humano e câmera fotográfica	6
2.5	Fotodiodo e conversão de luz em corrente elétrica	6
2.6	Matriz RGB	7
2.7	Etapas da visão computacional	8
2.8	Human-occupied Vehicle	11
2.9	Remoted Operated Vehicles	11
2.10	Autonomous Underwater Vehicle	11
2.11	Glider Skandi Explorer	12
2.12	Visão humana x visão canina	14
2.13	Detecção de alvos com uso de RADAR e Sonar	15
2.14	Tela de apresentação RADAR	15
2.15	Sistema Tatico - Siconta	16
2.16	Neuronio fonte:(ROCHA, 2016)	17
2.17	Sinapse	17
2.18	Modelo foto-perceptron de Rosenblatt	19
2.19	Modelo de Perceptron não - linear	19
2.20	Exemplo perceptron - linear	20
2.21	Funções de ativação suave	21
2.22	Função degrau unitário	22
2.23	Principais funções de ativação	23
2.24	Principais arquiteturas de RNAs fonte:(VENN, 2016)	24
2.25	A função XOR: uma função linearmente inseparavel fonte: autor	25
2.26	Portas AND OR XOR fonte:autor	25
2.27	MLP: solução porta XOR fonte:(LAVRENKO, 2016)	26
2.28	MultiLayer Perceptron - Arquitetura fonte:(RAMCHOUN <i>et al.</i> , 2016)	28
2.29	MultiLayer Perceptron - Retropropagação do erro fonte:(JANA, 2019)	29
2.30	Taxas de aprendizagem e função Loss fonte: < https://cs231n.github.io/neural-networks-3/ >	30
2.31	Melhores curvas de aprendizagem fonte:(HAYKIN, 2001)	31
2.32	Hiperparâmetros em gupos fonte:autor	32
2.33	Entrada do valor 0 - MNIST fonte:(DEEPLIZARD..., 2021)	34
2.34	Menu DEEPLIZARD Demo fonte:(DEEPLIZARD..., 2021)	34
2.35	DEEPLIZARD Demo- filtros fonte:(DEEPLIZARD..., 2021)	34
2.36	Produto Escalar - filtro fonte:(DEEPLIZARD..., 2021)	35
2.37	Saídas de filtros de convolução fonte:(DEEPLIZARD..., 2021)	36
2.38	Detecção de objeto com R-CNN fonte:(GIRSHICK <i>et al.</i> , 2014)	37
2.39	Detecção de objeto com YOLO fonte:(REDMON <i>et al.</i> , 2016)	38
2.40	Cabos subamrinos no mundo fonte:< https://www.submarinecablemap.com/ >	40
2.41	Recursos e atividades econômicas na Amazônia Azul fonte:(PADILHA, 2020)	42
2.42	Achado arqueológico em Thonis-Heracleion fonte:(MALVA, 2020)	43
2.43	Igreja submersa Sagrado Coração - Petrolândia	44

2.44	Áreas SAR- Marinha do Brasil fonte: < https://www.marinha.mil.br/dgn/?q=salvamar >	44
2.45	Tipo de minas subaquáticas fonte:(TRUVER, 2012)	46
3.1	Amostra de imagens do dataset SEABED-KLSG-II fonte: SeaBed-KLSG-II	49
3.2	Amostra de imagens do dataset MARIS-Project-Set1 fonte: MARIS-Project 49	
3.3	Amostra de imagens do dataset UOT-100 fonte: UOT-100	50
3.4	Etapas envolvidas na criação de modelo computacional.	50
4.1	Arquitetura da rede neural utilizada.	54
4.2	Curvas de aprendizagem médias para $\eta = 0.01$	54
4.3	Curvas de aprendizagem médias para $\eta = 0.1$	55
4.4	Curvas de aprendizagem médias para $\eta = 0.5$	55
4.5	Curvas de aprendizagem médias para $\eta = 0.9$	55
4.6	Melhores curvas de aprendizagem selecionadas.	56

Lista de Siglas

ATR	<i>Automatic Target Recognition</i>
AUV	<i>Autonomous Unmanned Vehicle</i>
CMM	<i>Contra Medidas de Minagem</i>
COC	<i>Centro de Operações de Combate</i>
DL	<i>Deep Learning</i>
HOV	<i>Human Occupied Vehicles</i>
IA	<i>Inteligência Artificial</i>
ML	<i>Machine Learning</i>
OTE	<i>Occipitotemporal</i>
PF	<i>Pré-frontal</i>
RNA	<i>Rede Neural Artificial</i>
ROV	<i>Remoted Operated Vehicle</i>
SAS	<i>Synthetic Aperture Sonar</i>
TI	<i>Temporal Inferior</i>
TL	<i>Transfer Learning</i>
TSP	<i>Temporal Superior Polissensorial</i>
ZEE	<i>Zona Econômica Exclusiva</i>

Introdução

1.1 Definição do Problema

Em um planeta cuja superfície é composta por 71% de água, por diversas vezes, as águas se mostraram ser o palco de grandes eventos da história. Desde a explosão das grandes navegações até os dias de hoje, os mares são responsáveis por conectar nações de todo o mundo.

Entretanto, esses mesmos mares que são diariamente navegados em suas superfícies, ainda escondem muitas coisas em suas profundezas. Diversas foram as tentativas de adentrar nas suas maiores profundidades pelo homem, porém o desafio sempre foi muito grande. As pressões muito grandes e seus efeitos na fisiologia do corpo humano, assim como a incapacidade do homem respirar na água delimitam as possibilidades de um mergulhador. Equipamentos para levar pessoas para profundidades maiores foram criados para superar essa dificuldade, mas sempre envolvendo riscos aos tripulantes.

A partir de navios submersíveis, foram desenvolvidos os primeiros navios submarinos, uma tecnologia disruptiva, porém desenvolvida fundamentalmente para emprego militar. Os navios submarinos, ou simplesmente submarinos, deram ao homem a capacidade de navegar com certa liberdade em águas profundas, porém, com a capacidade de "enxergar" e explorar o fundo do mar de forma bem limitada. As águas profundas guardavam novas surpresas para o homem: a luz apresenta um comportamento bem diferente de quando está no ar sofrendo muita refração, as ondas eletromagnéticas sofrem muita atenuação, e as variações de salinidade e temperatura provocam efeitos na velocidade das ondas acústicas.

A detecção de objetos naturais ou artificiais abaixo d'água é uma capacidade desejável em diversas áreas de conhecimento, desde pesquisas na área biológica e de exploração do oceano (ROSLI *et al.*, 2021), na indústria de petróleo e gás (NADIMI *et al.*, 2021), exploração de artefatos arqueológicos no fundo do mar (MORONI *et al.*, 2014), prospecção de minérios em baixo d'água (SIDDHARTHA *et al.*, 2021) dentre outros. A capacidade de detectar e reconhecer um objeto de interesse a partir da captura de informações do ambiente pode ser considerado um ponto crucial para a exploração.

Em contraste aos diversos reveses sofridos pelo homem, o aprimoramento dos dispositivos de sensoriamento, em conjunto com a explosão do aumento de pesquisas de Inteligência Artificial (IA) corroboram para um cenário mais favorável. O surgimento dos veículos

não-tripulados com múltiplos sensores podem superar muitas das dificuldades impostas pela natureza submarina. Entretanto, essas ferramentas em conjunto só poderão ser úteis se houver por trás de seu funcionamento um algoritmo capaz de processar e reconhecer objetos de interesse, em meio às imagens geradas pelos diversos sensores.

A IA tem contribuído bastante para a resolução de problemas relacionados a detecção, classificação e reconhecimento de objetos. O trinômio veículo - sensor- processamento vem delineando um novo escopo, a partir de novas descobertas em cada um de seus aspectos, seja na melhoria dos veículos, na possibilidade de inserir neles sensores múltiplos, ou até mesmo na evolução dos algoritmos relacionados à visão computacional. Dessa forma, a convergência de tecnologias e técnicas recentes podem trazer à tona uma nova era para a exploração do fundo dos mares.

1.2 Objetivos

Geral: Desenvolver um modelo computacional, baseado em aprendizagem de máquina, capaz de reconhecer objetos naturais e artificiais no ambiente subaquático.

Específico:

- Montar um dataset de imagens subaquáticas rotuladas
- Desenvolver uma rede do tipo Multi layer Perceptron (MLP) para aprendizagem
- Avaliar resultados da MLP
- Publicar trabalhos acadêmicos dos resultados obtidos

1.3 Justificativa

O presente estudo se justifica pela importância que as águas jurisdicionais brasileiras, sobretudo a Zona Econômica Exclusiva (ZEE) representam para o país no desenvolvimento econômico e social, nas mais diversas atividades de exploração petróleo e gás, prospecção e mapeamento de recursos minerais, atividades pesqueiras, monitoramento de infraestruturas críticas no mar, busca e salvamento no mar, dentre outros.

Segundo relata o Almirante de Esquadra Ilques Barbosa Junior em "No mar, a riqueza e o futuro das nações":

Dos mares retiramos cerca de 95% do petróleo, 80% do gás natural e 45% do pescado produzidos no País. Pelas rotas marítimas escoamos mais de 95% do comércio exterior brasileiro. No mundo globalizado, cada vez mais dinâmico e conectado em que vivemos, é por meio de cabos submarinos que flui o tráfego de dados responsável por quase toda a comunicação com o mundo, incluindo a internet.([ESQUADRA, 2019](#))

1.4 Organização da dissertação

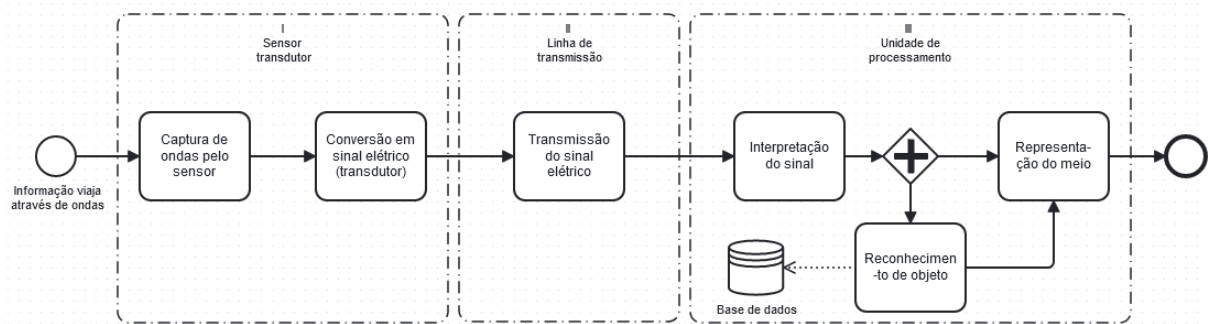
O trabalho se encontra organizado das seguinte maneira:

1. O capítulo [1](#) contém a introdução e contextualização do trabalho;
2. O capítulo [2](#) engloba toda a revisão de literatura abrangendo o conhecimento a cerca dos sensores e veículos, ebm como os métodos de detecção e uso de IA para esse fim desde o seu início;
3. O capítulo [3](#) fará uma descrição dos recursos e métodos utilizados para montagem do dataset de imagens subaquáticas utilizado, bem como e criação e simulação do modelo computacional; e
4. O capítulo [4](#) é reservado para a apresentação e discursão dos resultados obtidos pela simulação.

Revisão da Literatura

2.1 Representação e reconhecimento de objetos: um modelo

Figura 2.1: Diagrama simplificado: processo de representação do ambiente

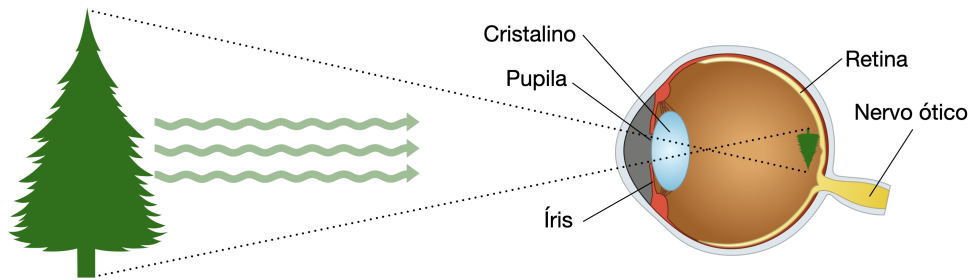


Fonte: Autor.

A detecção, classificação e reconhecimento de objetos abaixo d'água são tarefas que requerem um sistema capaz de gerar uma representação do ambiente subaquático que seja inteligível para o usuário para posterior análise. A detecção de um objeto consiste em perceber algum objeto de interesse no ambiente, a classificação é a capacidade de atribuir ao objeto detectado uma classe, e o reconhecimento consiste em identificar um objeto conhecido. O sistema capaz de gerar uma representação de um ambiente deve conter pelo menos três elementos: o veículo, o sensor, e o centro de processamento. O sensor é o elemento responsável por capturar informação do ambiente, o veículo é um elemento que possui mobilidade e é capaz de transportar um ou mais sensores, e o centro de processamento é o elemento que interpreta a informação capturada pelo sensor (SILVEIRA *et al.*, 2022). Um quarto elemento poderia ser citado como transdutor, como aquele que converte a informação capturada em sinal a ser interpretado, mas para efeito de simplificação, pode-se considerar que este está acoplado ao sensor.

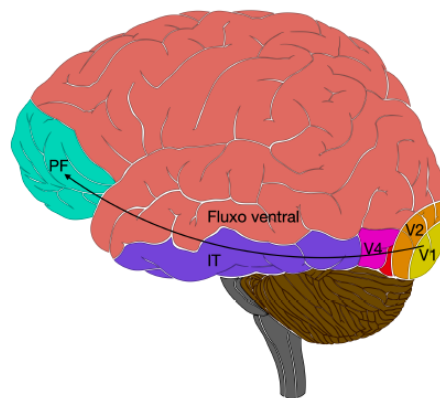
A Figura 2.1 mostra o processo pelo qual o sistema citado realiza a representação do ambiente. O processo se inicia com a captura de informação. O sensor captura informação do ambiente que é transportada por meio de ondas. A informação (onda) capturada pelo sensor deve então ser transformada em um sinal elétrico por um transdutor, para que seja possível transportar, processar e manipular essa informação. O sinal elétrico percorre a linha de transmissão até chegar a um centro de processamento, que será capaz de interpretar esse sinal e gerar uma representação do ambiente baseado nas informações capturadas. O centro de processamento poderá, também, ser treinado para reconhecer objetos de interesse. Na seção 2.2 será apresentado um exemplo prático deste processo.

Figura 2.2: Visão Humana: Captura de informação do ambiente



Fonte: Autor

Figura 2.3: Visão humana: fluxo ventral



Fonte: Autor

2.2 Visão humana: um exemplo de modelo de reconhecimento de objeto

Para entender melhor a dinâmica de funcionamento do modelo representado na Figura 2.1, considere a seguinte analogia com a visão humana: os olhos são sensores, os fotorreceptores na retina são transdutores, o nervo óptico é a linha de transmissão e o cérebro é o centro de processamento.

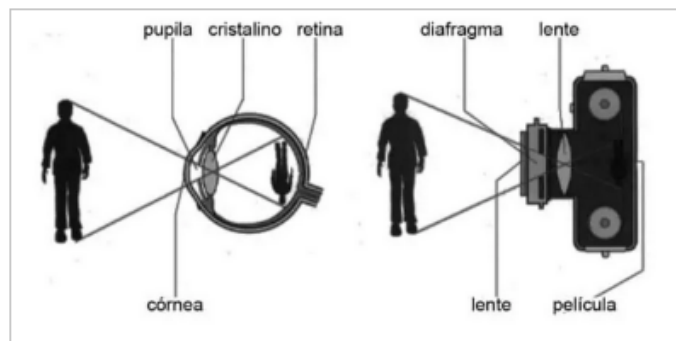
A Figura 2.2 ilustra o processo de captura de informação do ambiente pelo olho humano. A luz refletida de um objeto atravessa o cristalino que aprisiona o estímulo da onda de luz na retina. A retina transforma padrões de luz projetados em fotorreceptores em sinais neurais que são transmitidos através do nervo óptico até centros visuais especializados no encéfalo, gerando padrões dinâmicos e complexos de luz de intensidade e cor variáveis.

O sistema visual dos seres humanos é amplamente organizado em duas vias de processamento segregadas, uma corrente ventral para visão de objetos e uma corrente dorsal para visão espacial (KONEN; KASTNER, 2008). O fluxo ventral da informação sensorial visual compreende os córtices V1, V2, V4, occipitotemporal - OTE, temporal inferior - TI, temporal superior polissensorial -TSP e pré-frontal - PF (Figura 2.3). Os córtices

TI e OTE fazem interface com o hipocampo para registro da memória de longo prazo e com o córtex PF para processamento da informação na memória de trabalho. Quando a informação que foi capturada do ambiente chega ao córtex PF, pode-se dizer que o cérebro(unidade processadora) gera a visualização do ambiente, ou seja, é neste momento que o indivíduo visualiza o objeto capturado demonstrado na Figura 2.2. O reconhecimento de um objeto, no caso da visão humana, envolve uma combinação ou comparação de informações sensoriais que chegam com informações armazenadas(LIEBE *et al.*, 2012).

2.3 Sistema óptico artificial, imagem digital e visão computacional

Figura 2.4: Olho humano e câmera fotográfica

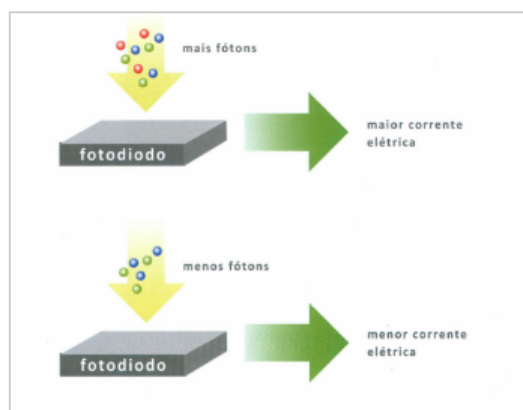


Fonte:

<https://blog.emania.com.br/olhos-vs-camera-fotografica-comparacao-do-olho-humano-a-uma-camera-digital/>

Assim como no caso da visão humana explicitado na seção 2.2, um sistema artificial pode ser criado para representar uma imagem conforme o diagrama da figura 2.1. De fato, como aponta Barelli (2018): “ A visão computacional também pode ser vista como um complemento da visão biológica”. No caso de um sistema artificial óptico, por sua vez, é possível se observar várias semelhanças com o sistema da visão humana.

Figura 2.5: Fotodiodo e conversão de luz em corrente elétrica

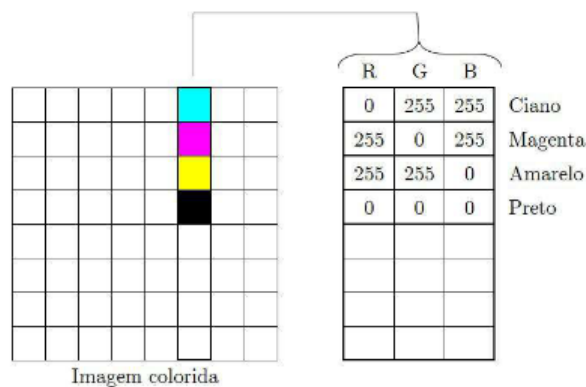


Fonte: (PALERMO, 2015)

Uma câmera fotográfica de filme faz a captura das ondas de luz do ambiente, passando em um primeiro momento por meio de uma lente externa que faz papel semelhante à córnea. De maneira análoga ao olho humano, a câmera possui mecanismos de ajuste de foco e controle de luz tendo como resultado a geração de imagens invertidas em uma superfície sensível à luz (Figura 2.4). Enquanto a íris pode dilatar ou contrair regulando o tamanho da pupila controlando assim, a quantidade de luz capturada, em movimento semelhante, o obturador de uma câmera controla a exposição da superfície sensível à luz. A lente interna possui um mecanismo de ajuste de foco semelhante ao cristalino e a película de um filme fotográfico, por sua vez, é sensibilizada pela luz, semelhantemente ao que ocorre na retina.

Assim como o olho humano e a câmera de filme, uma câmera digital também possuirá os mecanismos de captura e controle de passagem de luz. Neste caso, a informação da luz é recebida por um sensor eletrônico(fotodiodo) que, a depender da quantidade de fótons, gera uma determinada corrente elétrica codificando assim a informação da luz capturada (Figura 2.5). Uma matriz de fotodiodos faz o papel do transdutor e a informação que agora está contida em uma corrente elétrica pode ser armazenada ou enviada para um circuito de processamento. Se os fotodiodos forem arranjados de maneira que se conheça a posição de cada um deles, e for possível se fazer a leitura de luminosidade em cada ponto, é possível se reconstruir a imagem que incidiu nessa matriz de diodos(PALERMO, 2015).

Figura 2.6: Matriz RGB



Fonte: (GONZALEZ, 2010)

As imagens digitais são representações formadas a partir das informações armazenadas ou capturadas por um dispositivo eletrônico, onde as informações são geradas por meio de matrizes. Como os circuitos digitais interpretam valores binários, se cada elemento da matriz contiver um byte de informação (8 bits ou 2^8 valores), cada elemento da matriz poderá assumir um valor decimal entre 0 e 255 (256 valores).

Com isso, a imagem capturada pode ser recuperada(montada) por meio de uma composi-

ção de pontos elementares da matriz, conhecidos como pixels¹. Um método muito usado para atribuição de valores de cor a um pixel foi proposto por GONZALEZ (2010), onde todas as cores apresentadas são uma combinação de três cores elementares: o vermelho, o verde e o azul. A partir desse princípio, é possível se constituir uma imagem onde, na verdade, cada pixel é um valor oriundo da combinação dos valores do vermelho(R), verde(G) e azul(B). É desta forma que se compõe uma escala RGB (2.6) e se constrói uma imagem a partir da informação que foi capturada por um dispositivo eletrônico.

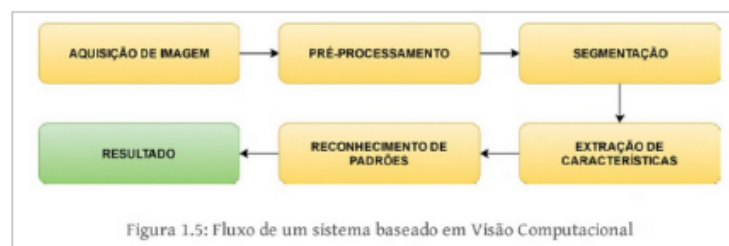
Diferentes dispositivos de captura deverão produzir resultados diferentes, devendo-se levar em consideração o tipo de emprego para a escolha do melhor dispositivo. De acordo com Shapiro & Stockman (2000): "Diferentes tipos de radiação que refletem ou penetram em objetos no mundo físico, podem ser ‘percebidos’ por diferentes dispositivos de imagem".

Uma outra questão envolvendo a captura é a calibração de lentes que permite determinar as características geométricas e radiométricas do sensor. Piteri & Rodrigues (2011) aponta os principais pontos-chave que devem ser considerados para a questão da calibração das câmeras como:

- os modelos de câmeras e sistemas de referência envolvidos;
- os modelos matemáticos para correção de erros;
- o uso de modelos estocásticos para estimativas de parâmetros;
- o método de calibração

As várias combinações possíveis desses pontos-chave irão levar a diferentes resultados e, por isso, eles devem ser levados em consideração para a melhor configuração de captura para cada emprego.

Figura 2.7: Etapas da visão computacional



Fonte: (BARELLI, 2018)

A visão computacional estuda e implementa sistemas capazes de enxergar por meio de processos artificiais, implementados por hardwares e softwares (BARELLI, 2018). Ainda

¹A palavra pixel se origina do inglês *picture+element*, e se refere ao menor elemento capaz de representar uma cor

segundo [Barelli \(2018\)](#), a visão computacional ocorre em etapas conforme diagrama da Figura 2.7.

O exemplo da câmera digital descreveu como ocorreu a captura da imagem e codificação da informação para ser usada posteriormente. A partir daí, tem-se uma imagem digitalizada com características codificadas. Uma fase de pré-processamento promove o realce de bordas e formas geométricas dos objetos de interesse contidos na imagem. A partir desse resultado, os objetos de interesse realçados podem ser segmentados em várias imagens, para que seja possível a aplicação de extração de características em cada um deles.

O reconhecimento de padrões que ocorre a partir da extração de características de um determinado objeto é uma questão central deste trabalho. Há atualmente diversas abordagens de extração de características existentes para a execução do reconhecimento de padrões. De acordo com ([JIANG, 2009](#)), essas abordagens podem ser organizadas em quatro grandes categorias:

- métodos baseados em conhecimento humano;
- abordagens baseadas na estrutura local da imagem;
- técnicas baseadas em estrutura global; e
- abordagens estatísticas baseadas em aprendizagem de máquina

No decorrer deste trabalho, será dado foco ao desenvolvimento de sistemas ópticos artificiais de captura de imagens e método de extração e reconhecimento de padrões baseados em aprendizagem de máquina.

2.4 Detecção, representação e reconhecimento no ambiente subaquático

A exploração dos oceanos e do ambiente marinho necessita do uso de recursos capazes de capturar informações em locais de difícil acesso e representá-las da maneira mais fidedigna possível ao usuário, possibilitando assim, a execução de tarefas diversas que podem envolver alta complexidade, ou até mesmo risco ao pessoal envolvido. Conforme foi mencionado na seção 2.1, um sistema genérico e simplificado capaz de realizar esse tipo de tarefa precisa do veículo, do sensor e do centro de processamento. A seção 2.2 descreveu como a visão humana é capaz de gerar uma representação do ambiente, onde o corpo humano faz o papel do veículo, os olhos são os sensores, e unidade de processamento e

representação é o cérebro. Nesta seção, serão abordados os veículos, os sensores e maneira pela qual o sinal é processado em uma tarefa de representação e reconhecimento de objetos.

2.4.1 O veículo

Tradicionalmente, os veículos utilizados para exploração do ambiente marinho são navios de superfície ou veículos submersíveis que podem transportar sensores na própria estrutura ou em um dispositivo lançado na água. Conforme novas tecnologias foram surgindo, e devido à necessidade crescente de se reduzir a exposição de pessoal em tarefas que envolvem risco, começou-se a adotar o uso de veículos não-tripulados para o mesmo fim. Os veículos não-tripulados podem ser teleoperados remotamente ou ser autônomos. Quando operados remotamente são controlados por um navio-mãe ou base fixa em terra. Quando autônomos, são capazes de realizar tarefas tomando suas próprias decisões, com base em algoritmos de Inteligência Artificial (IA).

De acordo com [Cui \(2018\)](#), os principais veículos submersíveis utilizados na exploração do ambiente marinho são:

- *Human-occupied vehicles*(HOV), também conhecidos como *Manned submersible*(MS), isto é, submersível tripulado;
- *Remoted Operated Vehicles*(ROV), também conhecidos como veículos operados remotamente;
- *Autonomous Underwater Vehicles*(AUV), conhecidos como veículo autônomo subaquático;
- ROV e AUV híbridos; e
- *Underwater gliders*, ou simplesmente *gliders*, conhecidos como planadores subaquáticos.

O HOV é um recurso muito bom para a exploração e investigação do ambiente marinho, que por transportar pessoal especializado, possui grande autonomia e capacidade decisória. Porém, não é recomendável em tarefas de risco, como detecção de alvos perigosos.

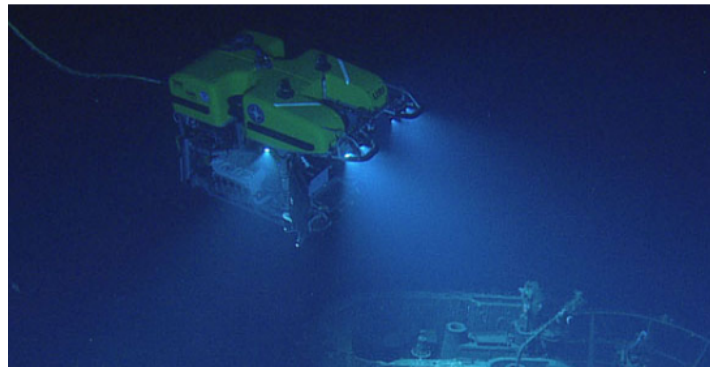
O ROV é um dispositivo amarrado por um cabo a um navio e controlado remotamente. Ele é adequado para a inspeção e exploração do ambiente marinho, possui risco reduzido em relação ao HOV, mas sua autonomia se limita pelo cabo que o liga ao navio. Esses veículos são normalmente equipados com câmeras e empregados em situações que requerem informação visual do local([GHAFOOR; NOH, 2019](#)).

Figura 2.8: Human-occupied Vehicle



fonte: (COMMERCE, 2001)

Figura 2.9: Remoted Operated Vehicles



fonte: (COMMERCE, 2001)

Figura 2.10: Autonomous Underwater Vehicle

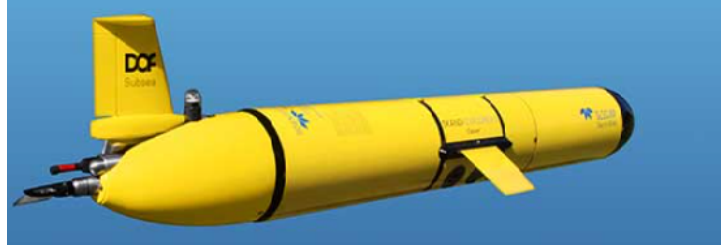


fonte: (COMMERCE, 2001)

O AUV é um veículo que conta com uma grande variedade de sensores, tanto acústico quanto visual, e são empregados para tomada de decisão inteligente (FABBRI *et al.*, 2018). Também podem ser especialmente úteis para monitoramento, observação e inspeção, uma vez que são auto-propulsados e podem executar muito bem essas tarefas por algumas

horas(GHAFOOR; NOH, 2019). Embora tenham autonomia razoável e anule o risco de acidentes pessoais, sua capacidade decisória depende do desenvolvimento de algoritmos avançados.

Figura 2.11: Glider Skandi Explorer



fonte: (SUBSEA, 2018)

Os *gliders* (Figura 2.11) são planadores submersíveis que têm a capacidade de ficar no oceano por longos períodos com menos consumo de energia quando comparado aos outros veículos, porém ele é mais lento e mais difícil de controlar(GHAFOOR; NOH, 2019).

Além dos veículos submersíveis, há a possibilidade de usar veículos de superfície transportando sensores que podem estar presos em sua estrutura (casco) ou em um dispositivo rebocado. Lee *et al.* (2019) fez uso de um navio de superfície não - tripulado, também conhecido como *Unmanned Surface Vehicle*(USV) para a aquisição de imagens de sonar multifeixe em água turbulenta.

Cada veículo citado tem a sua particularidade e suas vantagens e desvantagens. Entretanto, é fato que há uma tendência a se adotar cada vez mais os veículos não-tripulados, especialmente aqueles capazes de transportar sensores diversos.

2.4.2 O sensor e o ambiente subaquático

Na seção 2.2 foi demonstrado como uma onda (nesse caso luz) é capturada pelo olho humano, e uma vez aprisionada no globo ocular e incidindo na retina, sensibiliza células especiais, que transformam a informação em um sinal elétrico preservando a informação.

A comunicação abaixo d'água por meio ótico sem fio, por exemplo, está propensa a efeitos severos de absorção e de espalhamento devido à natureza da luz visível quando ela se propaga no meio aquático (HASSAN *et al.*, 2022). No caso da detecção de um alvo, o efeito de espalhamento provoca distorções nas informações. Além disso, há a complexidade do ambiente aquático que possui correntes que variam no tempo, iluminação desigual no ambiente e turbulência da água que dificultam a captura de informação nesse ambiente(TENG; ZHAO, 2020).

Fatores como refração e absorção de luz, partículas suspensas na água e distorção de cor afetam a qualidade do dado visual, resultando em ruído e imagens distorcidas, reduzindo a performance da detecção de objetos abaixo d'água quando usando AUVs equipados com sensores óticos (FABBRI *et al.*, 2018).

Durante a Primeira Guerra Mundial, a marinha inglesa desenvolveu um equipamento pioneiro para fazer frente aos temidos submarinos alemães, os *U-boats*. Nomeado de ASDIC (*Anti-Submarine Detection Investigation Committee*), o equipamento que é considerado uma versão primitiva do SONAR (*SOund Navigation And Ranging*), foi desenvolvido para estimar a posição aproximada do submarino inimigo enviando pulsos capazes de ecoar no casco dos submarinos (HOOKER, 2021). Embora num primeiro momento essa tecnologia não tenha sido muito eficaz, melhorias feitas no SONAR deixariam um marco no desenvolvimento de tecnologias de detecção na água por meio acústico.

Com base nos princípios da acústica submarina, vários dispositivos semelhantes ao SONAR foram desenvolvidos. Estudos nessa área têm colaborado para a implementação do SONAR para exame de ativos marítimos, reconhecimento de alvos, e estudo e mensuração de fontes acústicas (NEUPANE; SEOK, 2020).

De acordo com Sun *et al.* (2021), de maneira geral, os sensores abaixo d'água são usados da seguinte forma:

- Sensores acústicos são usados para mapeamento do solo aquático, navegação de submersível e detecção de objeto;
- Sensores ópticos são empregados para exploração dos oceanos, inspeção de objetos, espectrofotometria e fluorofotometria para monitoramento de parâmetros ambientais; e
- Sensores eletromagnéticos usados para detecção de metais, recursos minerais e inspeção de cabos e dutos subaquáticos

É importante ressaltar que o uso de múltiplos sensores pode favorecer o refinamento da leitura do ambiente, uma vez que a incerteza de um sensor pode ser compensada pela leitura do outro em uma fusão de sensores.

2.4.3 Processamento do sinal, interpretação e reconhecimento

O correto processamento do sinal elétrico, que contém informações condensadas e codificadas do ambiente, é essencial para a consolidação da representação e do reconhecimento

de um objeto. Ao longo de cada etapa do processo demonstrado na Figura 2.1, idealmente, a informação se mantém íntegra, até que ela seja processada de maneira a gerar uma representação muito próxima da realidade.

Entretanto, na prática, perdas podem ocorrer em todas as etapas e até mesmo antes da captura da informação. As ondas que viajam no meio transportando informação podem sofrer atenuação, refração, interferência e serem combinadas com ruídos do ambiente.

A sensibilidade do sensor pode não ser adequada para a situação favorecendo novas perdas. A conversão de um sinal seja acústico, seja eletromagnético, em sinal elétrico nunca é perfeita e esta contribui para novas perdas na informação original.

Figura 2.12: Visão humana x visão canina



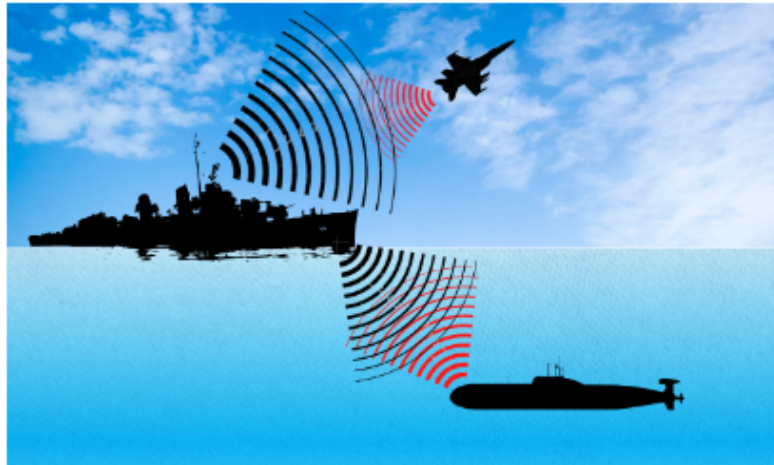
Fonte: (FERNANDO, 2018)

É importante notar que dois sistemas diferentes, ainda que não tenham nenhum tipo de problema em nenhuma das etapas do processo, devem gerar visualizações diferentes de um mesmo ambiente. Um exemplo simples disso é a comparação de dois sistemas visuais: o homem e o do cachorro (Figura 2.12). Percebe-se que, para um mesmo ambiente, há uma diferença no resultado para os dois sistemas visuais. Essa diferença se deve, principalmente, à diferença de configuração dos fotorreceptores da retina do homem e do cachorro.

Sistemas artificiais podem ser criados para prover visualização de uma determinada informação de um ambiente. Considere o exemplo de um navio de guerra em uma missão de reconhecimento dotado de sistemas RADAR e SONAR (Figura 2.13). Um sistema RADAR simples deve possuir uma antena fazendo o papel de sensor com um transdutor acoplado, uma linha de transmissão e um equipamento para processamento e visualização que fica localizado no Centro de Operações de Combate(COC). O equipamento gera uma visualização semelhante à da Figura 2.14.

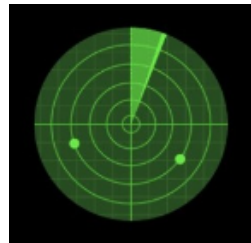
O navio da Figura 2.13 pode possuir um equipamento de Medidas de Apoio à Guerra Ele-

Figura 2.13: Detecção de alvos com uso de RADAR e Sonar



fonte: Autor.

Figura 2.14: Tela de apresentação RADAR



fonte: Autor.

Tabela 2.1: Alguns dos principais parâmetros de um RADAR

Nome do parâmetro	Descrição	Unidade
Frequência	Frequência do sinal	Mhz
Banda	Faixa a qual a frequência de emissão pertence	UHF , L , K etc
Largura do pulso (τ)	Tempo de duração de um pulso	micro segundos
Frequência de repetição de Pulso (FRP)	Número de pulsos em um determinado tempo	Pulsos por segundo
Comprimento de onda (λ)	Comprimento da onda de rádio transmitida	centímetros e
Largura de feixe (θ)	Largura angular do feixe de varredura do radar	graus ou radianos

Fonte: Autor.

trônica (MAGE), que consiste em uma antena receptora de grande sensibilidade, configurada para receber/capturar emissões de radares de outros navios, capturando importantes características desse sinal(ver tabela 2.1).

Os sinais capturados são enviados para o COC (Figura 2.15) para um equipamento capaz de realizar análise das características dos sinais, e por comparação com um banco de dados, pode ser possível identificar qual equipamento radar está emitindo aquele sinal.

Figura 2.15: Sistema Tatico - Siconta



Fonte: <<https://www.marinha.mil.br/emgepron/pt-br/sistema-de-controle-tatico-siconta>>

Essa informação poderá ser integrada ao sistema tático e ser correlacionada com um dos pontos-alvos mostrados no RADAR ampliando a visualização do cenário tático. Um sistema SONAR funciona de maneira muito semelhante ao RADAR, sendo a principal diferença que ele usa ondas acústicas e foi feito para detecção abaixo d'água.

2.5 Reconhecimento e aprendizagem de máquina

Um método eficiente de reconhecimento é um passo fundamental para a implementação do reconhecimento de alvos/objetos de forma automática, isto é, para a execução de *Automatic Target Recognition* (ATR). O ATR é um assunto muito abordado ultimamente e um dos pilares do desenvolvimento de um sistema de reconhecimento de objeto (alvo). A combinação de um bom método de detecção/classificação com o uso de navios autônomos e de uma combinação de sensores adequada tem revolucionado as mais diversas tarefas de mapeamento e busca de objetos abaixo d'água.

Para realizar acompanhamento, detecção, classificação e ATR usando sinais sonar, muitos métodos de processamento de sinal clássicos, táticas baseadas em aprendizagem de máquina (ML) e abordagens em aprendizagem profunda (DL) têm sido previstos e aplicados (NEUPANE; SEOK, 2020).

A interpretação do sinal para efeito de classificação e reconhecimento em sistemas artificiais ou naturais está diretamente ligada à capacidade de se extrair parâmetros do sinal. O reconhecimento depende da capacidade de classificar um objeto, seja por meio da análise de parâmetros ou por meio da comparação com amostras de objetos semelhantes. O uso de múltiplos sensores, assim como no exemplo do indivíduo vendo o pinheiro ou dos sistemas táticos, poderá ajudar a acumular mais informações sobre um mesmo objeto, melhorando assim o refinamento do seu reconhecimento.

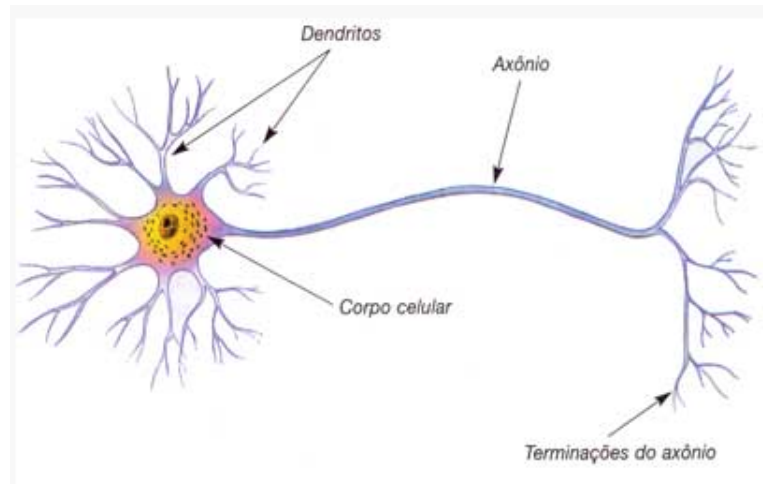
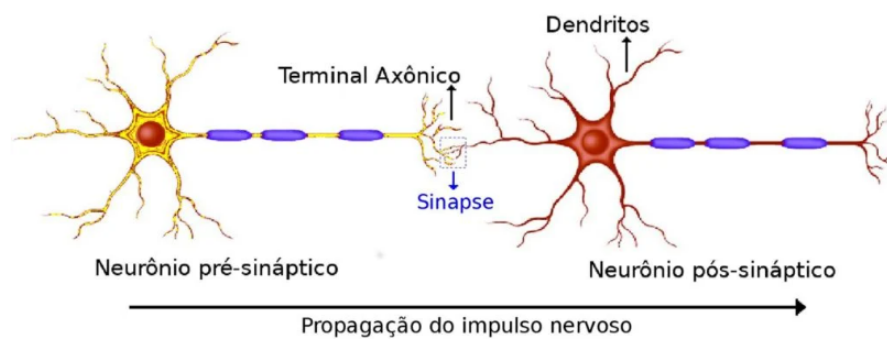


Figura 2.16: Neurônio
 fonte:(ROCHA, 2016)

Figura 2.17: Sinapse



Fonte: (FLORES, 2020)

Na seção 2.2 foi observado que no processo de reconhecimento de objeto pela visão humana, as áreas V4 e TI fazem uma conexão com estruturas do lobo parietal envolvidas no aprendizado e na memória, isto é, a capacidade de reconhecimento de um objeto está diretamente ligada ao nível de aprendizado e à memória daquele objeto. Para um sistema artificial, quando automatizado, isso não é diferente: para um sistema de reconhecimento de um determinado objeto reconhecer o objeto-alvo com acurácia, ele precisa de um bom banco de dados desse objeto. Portanto, a capacidade de um sistema, baseado em aprendizagem de máquina, de reconhecer um objeto dependerá do seu dataset(memória) e do seu treinamento (aprendizado).

2.6 As redes neurais e o perceptron

Uma rede neural tem como elemento básico um neurônio, que é uma célula especializada na transmissão de sinais elétricos contendo informação. O neurônio é composto por dendritos, corpo e axônio (Figura 2.16). O neurônio pode se ligar (conectar) a outros

neurônios por meio da sinpase (Figura 2.17), que é uma ligação formada entre terminal axonal de um neurônio com o dendrito de outro. Dessa forma, diversos neurônios podem se conectar formando uma rede neural (natural).

Na tentativa de se reproduzir artificialmente o cérebro humano (ou de mamíferos em geral), cientistas tentaram criar um modelo matemático do neurônio que pudesse ser simulado em um computador. Na busca por esse neurônio artificial, em 1943, Warren S. McCulloch e Walter Pitts publicaram o artigo "*A logical calculus of the ideas immanent in nervous activity*". Eles propuseram que, devido à característica "tudo-ou-nada" da atividade nervosa, os eventos neurais e suas relações poderiam ser tratados por meio de lógica proposicional (MCCULLOCH; PITTS, 1943). Uma conclusão relevante desse estudo foi a de que cada neurônio teria um *threshold*, isto é, um valor limite para uma excitação que deveria ser excedida para iniciar um impulso.

Em 1949, D.O. Hebb publicou o livro "*The Organization of Behavior: A NEUROPSYCHOLOGICAL THEORY*", onde ele tentou explicar como algumas memórias são instantaneamente estabelecidas e permanentes. Ele escreve o seguinte postulado:

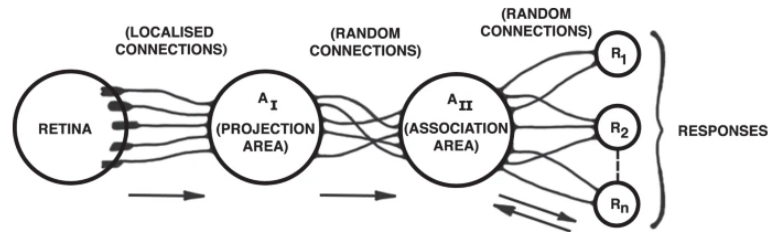
"Quando um axônio de uma célula A está próxima o suficiente de excitar a célula B e repetidamente ou persistentemente participa de sua ativação, algum processo de crescimento ou mudança metabólica ocorre em uma ou ambas as células, de tal forma que a eficiência de A, como uma das células de ativação de B, é aumentada"(HEBB, 2005)

O postulado de Hebb sugere que, quando há ativação de células simultâneas, há um fortalecimento de suas conexões, sendo esse um fator importante para a memória. Essa conclusão implica a necessidade de adaptação do modelo de McCulloch-Pitts, de modo que se deve considerar a atribuição de pesos aos estímulos de entrada, e não tão somente a característica de limiar de ativação.

A partir das descobertas de McCulloch-Pitts e de Hebb, Frank Rosenblatt modelou o que seria o primeiro *perceptron*. Ele publicou em 1958 o artigo "*The perceptron: a probabilistic model for information storage and organization in the brain*". Nesse trabalho, Rosenblatt ilustrou um modelo de um foto-perceptron (um *perceptron* que responde a estímulos de padrões ópticos). A Figura 2.18 mostra o diagrama original do foto-perceptron idealizado por Rosenblatt.

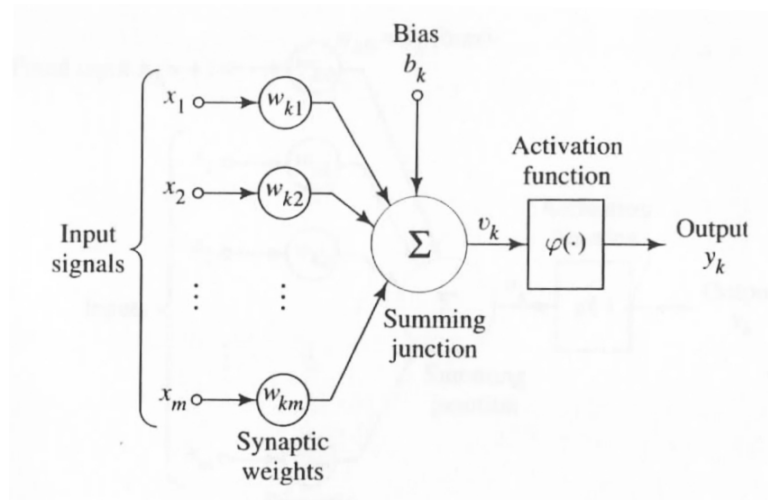
No modelo proposto, unidades sensoriais da primeira camada (retina) respondem a estímulos de acordo com o modelo "tudo-ou-nada". Essas unidades sensoriais estão localmente conectadas com as células da segunda camada (área de projeção). Entretanto,

Figura 2.18: Modelo foto-perceptron de Rosenblatt



fonte: (ROSENBLATT, 1958)

Figura 2.19: Modelo de Perceptron não - linear



fonte: (HAYKIN, 2001)

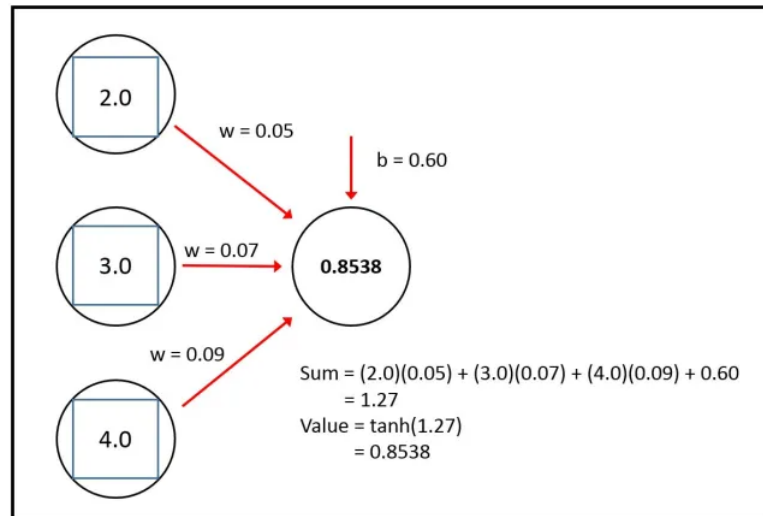
as conexões entre as camadas de projeção e associação, bem como entre as camadas de associação e resposta são aleatórias. Ao longo da rede, as transmissões das conexões são todas orientadas para frente (*feed-forward*), entretanto, no último trecho, isto é, entre as conexões da camada associativa e as *R-units* as conexões são estabelecidas em ambas as direções (ROSENBLATT, 1958).

A retina foi inspirada no sistema visual e um dos propósitos de Rosenblatt com o modelo proposto era desenvolver a capacidade de reconhecimento de padrões, capacidade essa que os humanos conseguem realizar sem dificuldade.

Embora esse modelo, na época, se mostrou capaz de ser treinado a reconhecer padrões de imagens e diferenciar, por exemplo, fotos de homens e mulheres, havia uma série de problemas do mundo real dos quais ele apresentava limitações. O perceptron não se mostrou capaz de tratar funções que não fossem linearmente separáveis, e como consequência não se mostrou capaz de realizar a função lógica XOR.

Haykin (2001) apresenta um modelo matemático de um neurônio (artificial) não-linear. A Figura 2.19 é uma reprodução do diagrama original encontrado no livro de Haykin.

Figura 2.20: Exemplo perceptron - linear



fonte:(MATOS, 2022)

De acordo com Haykin (2001), o modelo da Figura 2.19 descreve um neurônio k composto por vários elos de conexão (sinapses), cada um caracterizado por um peso ou força própria (w_{k1} , w_{k2} , etc). A junção aditiva soma as entradas ponderadas, caracterizando um combinador linear. A função de ativação, também conhecida como função restritiva, serve para limitar o intervalo permissível de amplitude do sinal a um valor finito. Um *bias* aplicado externamente, representado por b_k pode aumentar ou diminuir a entrada da função de ativação. Em termos matemáticos, o modelo pode ser escrito de acordo com a equação 2.1.

$$y_k = \phi(u_k + b_k)$$

onde:

$$u_k = \sum_{j=1}^m w_{kj}x_j \quad \text{soma de entradas ponderadas} \quad (2.1)$$

x_1, x_2, \dots, x_m : sinais de entrada

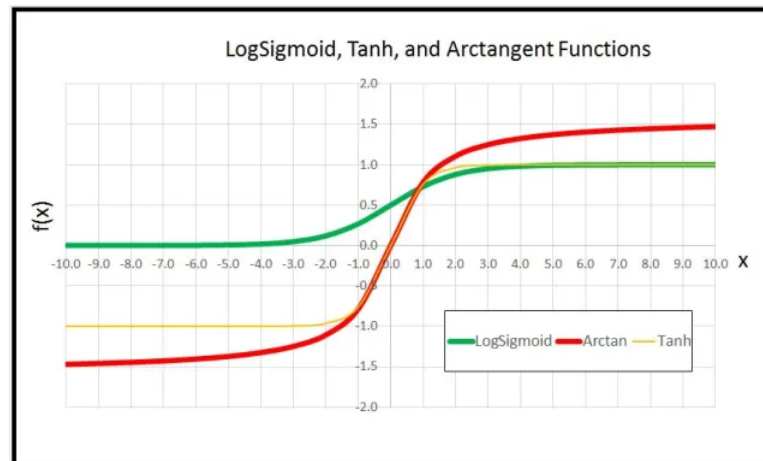
$w_{k1}, w_{k2}, \dots, w_{km}$: pesos sinápticos

b_k : bias

$\phi()$: função de ativação

A função de ativação $\phi()$ simula a condição de *threshold* do neurônio. Isso significa que, a depender da função de ativação, o resultado do somatório das entradas ponderadas com o bias pode gerar ou não a saída y_k . Se não houvesse a função $\phi()$ na equação 2.1, a saída y_k seria meramente uma combinação linear das entradas. Dessa forma, a função de ativação é o elemento que atribui a não-linearidade ao *perceptron*.

Figura 2.21: Funções de ativação suave



fonte:(MATOS, 2022)

A escolha da função de ativação deve ser feita levando-se em consideração a natureza do problema e o tipo de saída que se deseja, ou melhor o *range* de saída. Para entender melhor como isso funciona, considere um perceptron com os seguintes dados:

1. Entradas:

- $x_1 = 2.0$
- $x_2 = 3.0$
- $x_3 = 4.0$

2. Pesos:

- $w_{k1} = 0.05$
- $w_{k2} = 0.07$
- $w_{k2} = 0.09$

3. $b_k = 0.6$

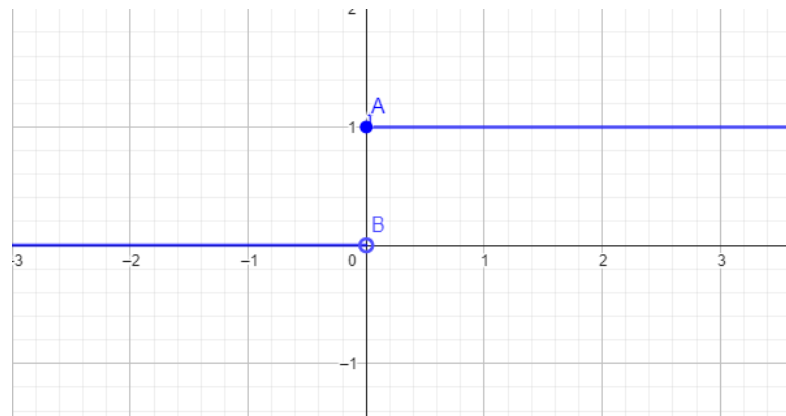
4. $\phi() = ?$

A Figura 2.20 mostra uma representação do *perceptron* com os valores enunciados. O valor da saída (y_k) é dada por:

$$y_k = \phi(\overbrace{(2.0 \times 0.05) + (3.0 \times 0.07) + (4.0 \times 0.09)}^{u_k} + \overbrace{0.6}^{b_k}) \quad (2.2)$$

$$y_k = \phi(1.27) \quad (2.3)$$

Figura 2.22: Função degrau unitário



fonte: Autor.

Se a função de ativação for a função tangente hiperbólica (\tanh), o valor da saída será

$$y_k = \tanh(1.27) = 0.85 \quad (2.4)$$

A Figura 2.21 mostra o gráfico das funções \tanh , \arctan e LogSigmoide . Pela observação das curvas, é possível perceber que o valor da saída y_k para a função \arctan será um pouco maior do que 0.85, enquanto que para a função LogSigmoide o valor será um pouco menor.

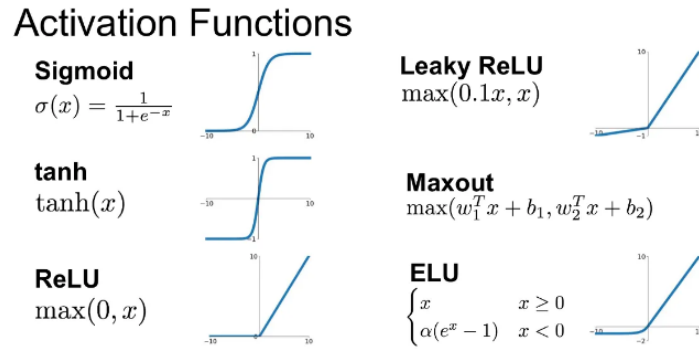
Se considerarmos agora a soma $u_k + b_k = -3$, as diferenças das saídas para cada função ficam maiores. Para a função \tanh , a saída será -1. A função \arctan retornará um valor de aproximadamente -1.25 e a função LogSigmoide retornará um valor próximo de zero.

Observando o comportamento para diferentes funções de ativação, é possível perceber a natureza não-linear do *perceptron*, bem como a diferença dos *ranges* de saída para cada $\phi()$. Portanto, a escolha da função de ativação deve atentar para o tipo de saída que se deseja e à natureza do problema proposto. As funções apresentadas na Figura 2.21 são funções suaves. Uma função mais abrupta poderia também ser utilizada como função de ativação.

$$u(t) \begin{cases} 1, & \forall t \geq 0 \\ 0, & \forall t < 0 \end{cases} \quad (2.5)$$

Considere a função degrau unitário dada pela equação 2.5. Nesse caso, o valor da saída para os valores 1.27 e -3 serão respectivamente 1 e 0. Observando a Figura 2.22, percebe-se que a função degrau vai retornar apenas valores 0 e 1 em todo o seu domínio.

Figura 2.23: Principais funções de ativação



fonte: (HIRA, 2021)

Além disso, a função degrau possui uma descontinuidade, conforme mostrado na equação 2.6.

$$\text{Limites de } u(t) \begin{cases} \lim_{t \rightarrow 0^-} u(t) = 0 \\ \lim_{t \rightarrow 0^+} u(t) = 1 \end{cases} \quad (2.6)$$

A função $U(t)$ pode ser usada quando se deseja uma saída do tipo 'sim ou não' 'ligado/desligado', 'verdadeiro/falso'. Outras funções, como a função tanh ou sigmoid podem restringir as saídas em um determinado range, porém de maneira bem mais suave do que na função degrau. A Figura 2.23 mostra um resumo das principais funções de ativação utilizadas na modelagem de perceptrons.

Um único perceptron pode sofrer limitações para solucionar problemas do cotidiano. Entretanto, é possível se criar um arranjo a partir da ligação de vários perceptrons, de forma que cada conexão de entrada do perceptron atual recebe o sinal de saída de um outro perceptron, originando formações mais complexas, conhecidas como Redes Neurais Artificiais (RNAs). A Figura 2.24 mostra um mapeamento das principais arquiteturas de RNAs utilizadas nos dias de hoje.

Conforme é possível ver, o próprio *perceptron* é considerado uma RNA elementar. Uma RNA, se for bem modelada e estiver aliada a uma boa capacidade computacional, poderá ser treinada em uma determinada tarefa para que ela retorne um determinado valor a um certo estímulo. Em outras palavras, pode-se dizer que a rede pode aprender com os estímulos do ambiente para posteriormente executar tarefas e solucionar problemas. Dessa forma, a RNA pode ser vista como um recurso de Inteligência Artificial (IA) capaz de realizar aprendizagem de máquina.

De acordo com Haykin (2001), a aprendizagem de máquina, no seu modelo simples, ocorre quando o ambiente fornece informação para o elemento de aprendizagem e partir daí, o elemento de aprendizagem pode aperfeiçoar a sua base de conhecimento.

As diversas arquiteturas de RNAs existentes nos dias de hoje permitem que elas sejam utilizadas para fins diversos. Entretanto, para este trabalho serão abordados apenas os principais recursos utilizados para tarefas de detecção, classificação e reconhecimento.

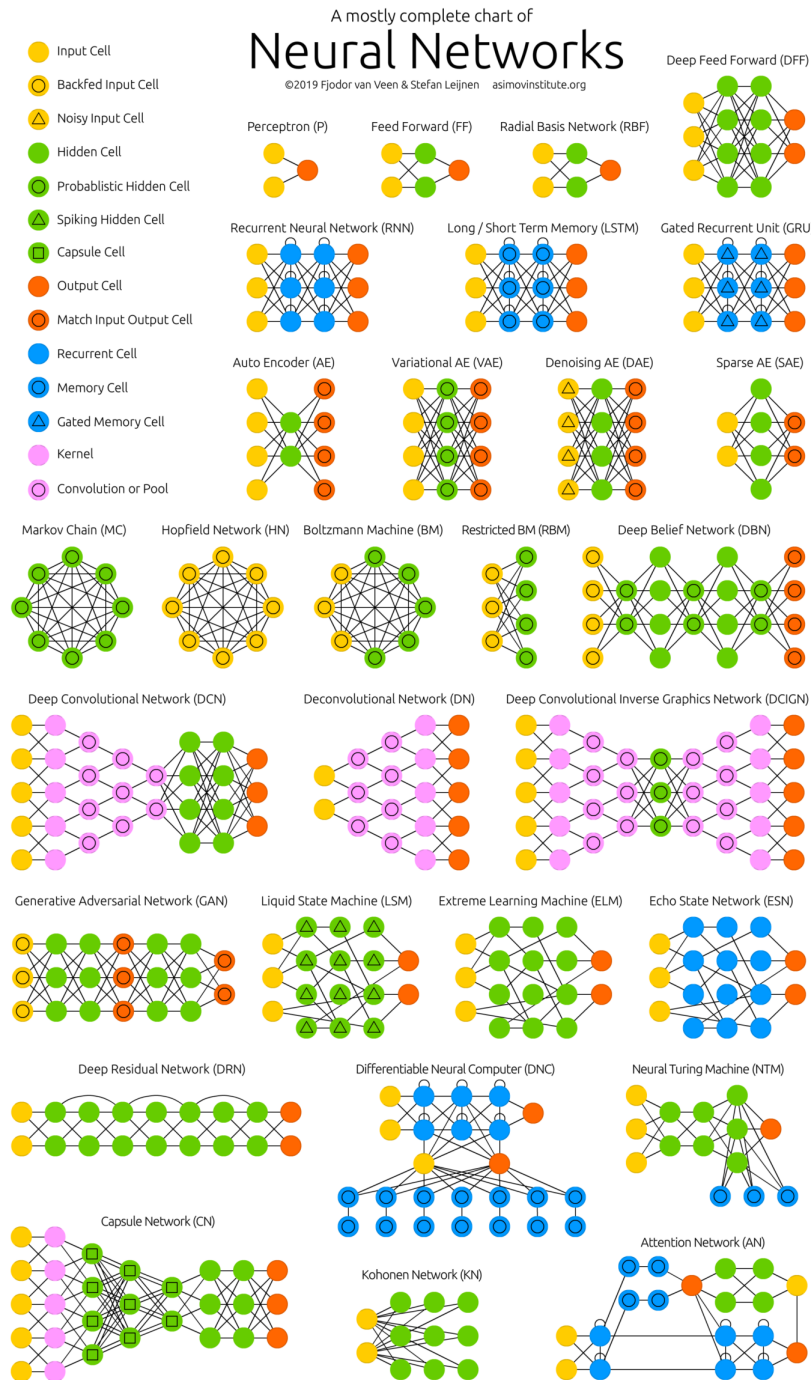


Figura 2.24: Principais arquiteturas de RNAs
fonte:(VENN, 2016)

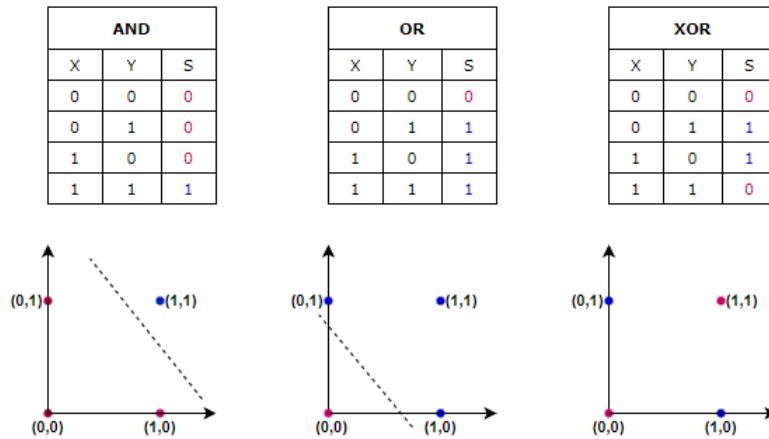


Figura 2.25: A função XOR: uma função linearmente inseparavel
 fonte: autor

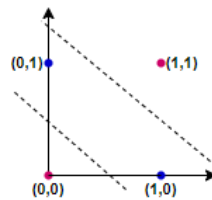


Figura 2.26: Portas AND OR XOR
 fonte:autor

2.7 A rede de perceptrons Multi camada

Conforme foi citado na seção 2.6, o *perceptron* apresenta dificuldades em trabalhar com funções que não são linearmente separáveis e um exemplo típico desse tipo de função é a função XOR. A Figura 2.25 mostra uma plotagem dos valores de saída das funções AND,OR e XOR, de acordo com suas tabelas-verdade. Percebe-se que, no caso das funções AND e OR é possível se traçar uma única reta que separa os possíveis valores de saída. Já no caso da função XOR isso não é possível.

Na Figura 2.26 é possível ver que é necessário pelo menos duas linhas para que seja possível separar as possíveis saídas, e com isso conclui-se que a função XOR não é linearmente separável, o que implica que um *perceptron* não é capaz de operar esse tipo de função. Essa característica do *perceptron* o torna um recurso limitado para tratar problemas reais do cotidiano.

Entretanto, é possível fazer uma combinação de *perceptrons* formando uma rede totalmente conectada (*fully connected*) onde esse problema pode ser facilmente contornável. A Figura 2.27 mostra uma representação em grafo de uma rede, capaz de resolver o problema da função XOR usando uma combinação de *perceptrons*.

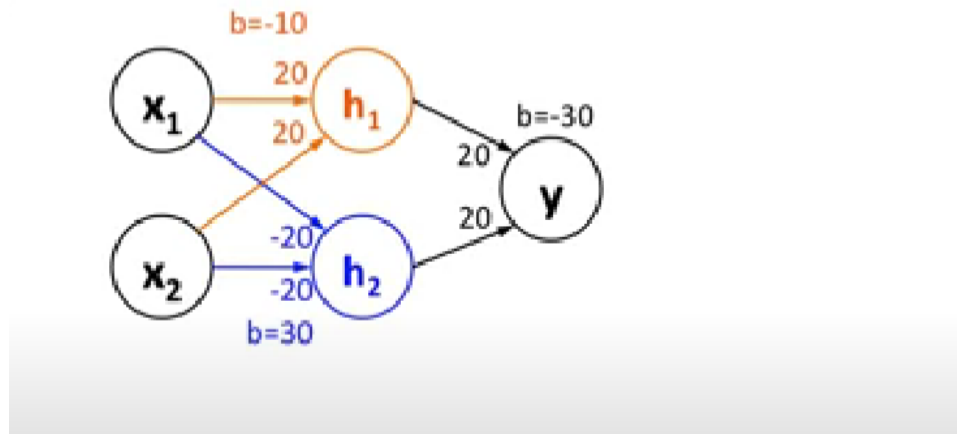


Figura 2.27: MLP: solução porta XOR
 fonte:(LAVRENKO, 2016)

X1	X2	Saida
0	0	0
0	1	1
1	0	1
1	1	0

Tabela 2.2: Tabela verdade da função XOR

Os dois primeiros nós dessa rede formam a camada de entrada, composta pelas entradas x_1 e x_2 , e portanto eles poderão assumir os valores os valores 1 ou 0. Os números apresentados nos arcos são os pesos. A segunda camada, composta pelos nós h_1 e h_2 , é conhecida por camada oculta (*hidden layer*) e os valores b são os bias. Por fim, y é a camada de saída composta por um nó. Portanto, percebe-se que essa rede possui três camadas: uma camada de entrada, uma camada oculta e uma camada de saída. Considere que para a rede apresentada, a função de ativação é a função sigmóide ($\sigma()$).

A partir da equação 2.1 é possível se desenvolver as equações para cada nó até se chegar no valor da saída y . Considere S_{21} a saída de h_1 e S_{22} a saída de h_2 . Assim, S_{21} , S_{22} e y podem ser escritas, respectivamente, como:

$$S_{21} = \sigma \times (20x_1 + 20x_2 - 10) \quad (2.7)$$

$$S_{22} = \sigma \times (-20x_1 - 20x_2 + 30) \quad (2.8)$$

$$y = \sigma \times (20h_1 + 20h_2 - 30) \quad (2.9)$$

A tabela verdade da função XOR é dada pela tabela 2.2. Ao aplicar os possíveis valores de entrada de x_1 e x_2 nas equações 2.7, 2.8 e 2.9 os seguintes resultados são obtidos.

$$(0,0) \begin{cases} S_{21} = \sigma \times (-10) = 0 \\ S_{22} = \sigma \times (30) = 1 \\ y = \sigma \times (20 - 30) = 0 \end{cases} \quad (2.10)$$

$$(0,1) \begin{cases} S_{21} = \sigma \times (20 - 10) = 1 \\ S_{22} = \sigma \times (-20 + 30) = 1 \\ y = \sigma \times (40 - 30) = 1 \end{cases} \quad (2.11)$$

$$(1,0) \begin{cases} S_{21} = \sigma \times (20 - 10) = 1 \\ S_{22} = \sigma \times (-20 + 30) = 1 \\ y = \sigma \times (40 - 30) = 1 \end{cases} \quad (2.12)$$

$$(1,1) \begin{cases} S_{21} = \sigma \times (40 - 10) = 1 \\ S_{22} = \sigma \times (-40 + 30) = 0 \\ y = \sigma \times (20 - 30) = 0 \end{cases} \quad (2.13)$$

Ao comparar os resultados obtidos com a tabela 2.2 percebe-se que a rede com os pesos atribuídos foi capaz de realizar a operação XOR. Com isso, concluímos que, ao criar uma rede com vários *perceptrons*, alguns problemas mais complexos que antes seriam impossíveis de se resolver passam a ser relativamente simples.

Esse tipo de arranjo, onde vários *perceptrons* se conectam formando uma rede com três camadas, sendo uma delas uma camada oculta, é conhecido como rede de perceptrons multi-camada, também conhecida como "*Multi Layer Perceptron*" (MLP). Uma MLP pode conter uma ou mais camadas ocultas e possui um alto grau de conectividade. No exemplo da Figura 2.27, podemos dizer que a rede é uma MLP contendo uma camada oculta, duas entradas, uma saída, e totalmente conectada.

Uma MLP pode ser usada para diversas tarefas a depender do seu arranjo, em termos de seus parâmetros, como por exemplo: quantidade de entradas, quantidade de saídas, quantidade de camadas ocultas, quantidade de neurônios das camadas ocultas.

A Figura 2.28 representa a arquitetura de uma MLP genérica contendo N camadas ocultas, com todas as conexões estabelecidas (*fully connected*) e distribuição de pesos variando entre as camadas. Para que seja possível projetar uma arquitetura adequada para uma

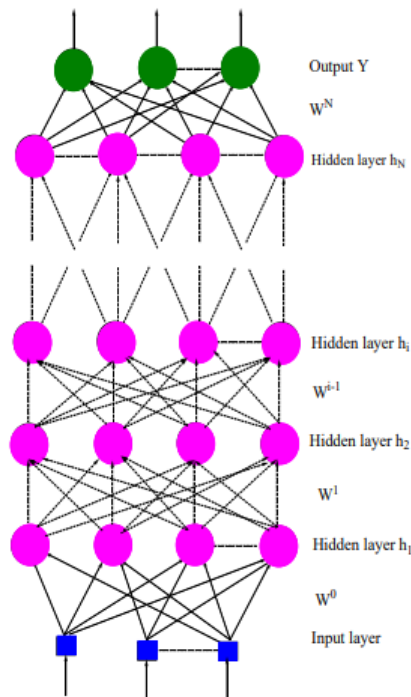


Figura 2.28: MultiLayer Perceptron - Arquitetura
 fonte:(RAMCHOUN *et al.*, 2016)

determinada questão, deve-se considerar o seguinte: a quantidade de entradas deve ser o número de *features* adotadas e o número de saídas deve ser o número de variáveis a serem preditas.

No exemplo da Figura 2.27, foi apresentada uma rede com pesos já distribuídas na camada oculta, para cada entrada de h_1 e h_2 e a rede já estava previamente configurada com uma camada oculta, dois neurônios na camada oculta e uma saída. Entretanto, o exemplo foi demonstrado apenas para comprovar que é possível se configurar uma MLP de modo que ela seja capaz de resolver o problema proposto, que nesse caso era a operação XOR.

Porém, mais importante do que saber que há possibilidade de resolver um determinado problema, é saber como projetar a arquitetura da MLP para esse fim, isto é, encontrar os hiperparâmetros da MLP.

O processo de aprendizagem de uma MLP serve para adaptar os pesos das conexões, a fim de se obter uma diferença mínima entre a saída da rede e a saída desejada(RAMCHOUN *et al.*, 2016). Dessa forma, conforme a rede treina ao longo das épocas (ciclos de processamento), ela vai ajustando os pesos, de modo que o erro dos resultados obtidos nas saídas sejam cada vez menores.

O mecanismo pelo qual a MLP faz ajuste dos pesos a cada época reduzindo o erro se chama retropropagação do erro (*error backpropagation*). Esse mecanismo funciona da seguinte maneira: ao comparar o valor da saída gerada pela rede em uma determinada

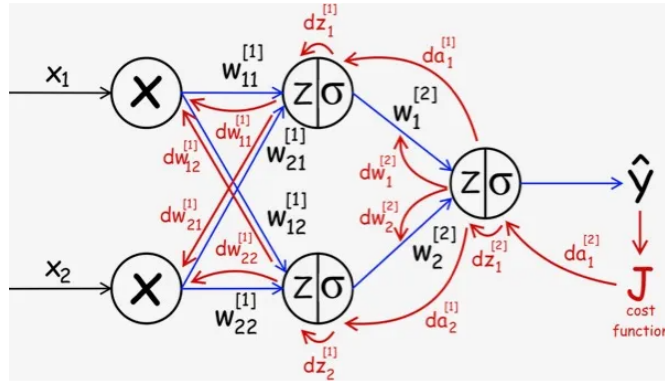


Figura 2.29: MultiLayer Perceptron - Retropropagação do erro
 fonte:(JANA, 2019)

época com o valor esperado, estima-se um valor de erro, que será usado pela rede para reajustar todos os pesos. O gradiente do erro em relação aos parâmetros da rede produz as alterações dos parâmetros de trás pra frente que mimizam o erro (Figura 2.29).

Quanto mais vezes a rede treina, isto é, quanto mais épocas, mais ajustes serão feitos e,consequentemente, menores serão os erros nas saídas e os pesos estarão melhor ajustados. Entretanto, é importante saber se, ao longo do treinamento, os valores dos pesos estão com tendência para a convergência. Ao longo das épocas, os valores dos pesos são ajustados conforme equação 2.14.

$$\omega_{novo} = \omega_{anterior} + \eta \times \overbrace{(d - y)}^{erro} \times x \tag{2.14}$$

onde,

- ω : pesos
 - d : saída desejada
 - y : saída
 - x : entrada
 - η : taxa de aprendizagem
- (2.15)

A rapidez e convergência da aprendizagem da rede pode ser controlada por meio de dois parâmetros: a taxa de aprendizagem e o momentum. Na equação 2.14, observa-se que erro na saída da reta está multiplicada pela variável η que é a taxa de aprendizagem.

A taxa de aprendizagem é uma variável que pode assumir valores reais entre 0 e 1 e que vai influenciar diretamente no 'passo' da atualização dos pesos. Isto significa que, quanto maior for o valor de η , mais rapidamente os pesos vão se atualizar, ou melhor, os pesos vão se atualizar a variações maiores a cada época. Deste modo, supõe-se que, quanto maior for o valor da taxa de aprendizagem, melhor é a rede pois ela tende a convergir mais

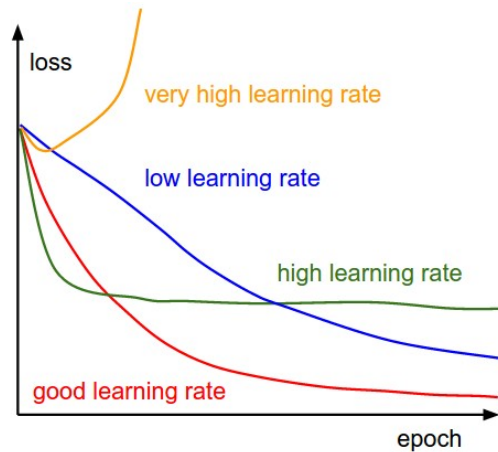


Figura 2.30: Taxas de aprendizagem e função Loss
 fonte: <<https://cs231n.github.io/neural-networks-3/>>

rapidamente. Entretanto, se a convergência for rápida demais, pode haver oscilação na rede, sendo neste caso mais interessante reduzir o η . No caso das curvas verde e laranja, por exemplo, o 'passo' é grande demais podendo ocorrer um *overshoot*, isto é, ultrapassar a marca do mínimo global, mesmo que haja um rápido decaimento de início. Já para a curva azul, um 'passo' pequeno demais, embora possa prevenir o *overshoot*, terá um grande custo computacional.

Na Figura 2.30 é possível ver o comportamento da função perda *loss* ao longo das épocas para diferentes taxas de aprendizagem. A linha vermelha mostra qual seria um bom valor de taxa de aprendizagem, no qual a função *loss* decai suavemente atingindo um valor muito baixo após uma certa quantidade de épocas. A linha azul mostra a evolução para um valor baixo de taxa de aprendizagem o que mostra que a convergência será muito lenta. A linha verde mostra que se a taxa for mais alta do que o adequado, a convergência será rápida, porém isso pode afetar negativamente a acurácia da rede. Por fim, a linha laranja mostra que se o valor for alto demais, isso pode afetar até mesmo a convergência.

Uma estratégia para se alcançar a convergência com rapidez e ao mesmo tempo tentar evitar os problemas das curvas verde e laranja da Figura 2.30, é usar um parâmetro que seja capaz de atualizar os pesos de acordo com a variação deste.

$$\omega_{novo} = \omega_{anterior} + \eta \times (d - y) \times x + \alpha(\omega_{anterior} - \omega_{anterior-1}) \quad (2.16)$$

Ao adicionar o termo $\alpha(\omega_{anterior} - \omega_{anterior-1})$ na equação 2.14, obtém-se a equação 2.16, onde α é a taxa de momentum e pode assumir valores reais no intervalo $[0, 1]$. Com isso, quando a variação dos dois pesos anteriores for grande, o novo termo com um valor de α adequado poderá acelerar a convergência. Se a diferença for pequena, o termo pouco

influenciará no ajuste do peso.

Haykin (2001) apresentou um experimento computacional, no qual realizou diversas simulações para diferentes valores de η e α em uma rede contendo dois neurônios ocultos. Os valores adotados para η e α , respectivamente foram:

$$\eta = \begin{cases} 0,01 \\ 0,1 \\ 0,5 \\ 0,9 \end{cases} \quad (2.17)$$

$$\alpha = \begin{cases} 0,0 \\ 0,1 \\ 0,5 \\ 0,9 \end{cases} \quad (2.18)$$

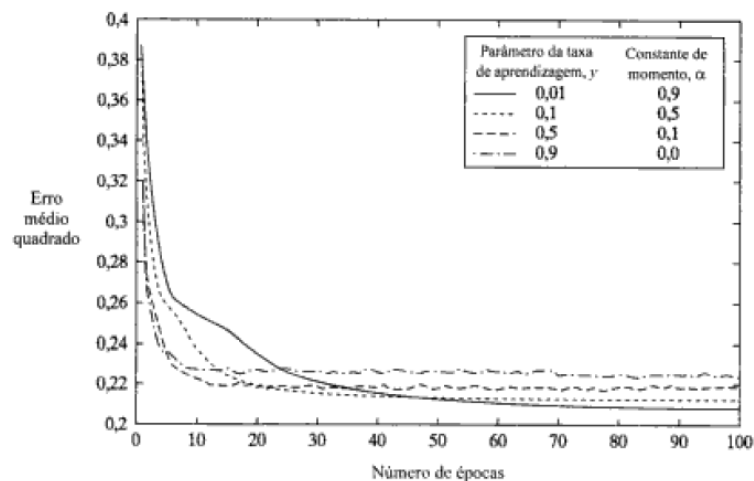


Figura 2.31: Melhores curvas de aprendizagem
fonte:(HAYKIN, 2001)

Onde η e α são valores reais pertencentes ao intervalo $[0,1]$. A equação 2.19 mostra o conjunto com todas as possíveis combinações de η e α .

$$P = \{(0.01, 0.0); (0.01, 0.1); (0.01, 0.5); (0.01, 0.9); (0.1, 0.0); (0.1, 0.1); (0.1, 0.5); (0.1, 0.9); (0.5, 0.0); (0.5, 0.1); (0.5, 0.5); (0.5, 0.9); (0.9, 0.0); (0.9, 0.1); (0.9, 0.5); (0.9, 0.9)\} \quad (2.19)$$

As melhores curvas de aprendizagem obtidas nas simulações pelas combinações destes valores podem ser vistas na Figura 2.31.

Outro parâmetro importante a ser considerado é o tamanho do lote de treinamento,

Hiperparâmetro	Valor
Número de entradas(x)	$x \in \mathbb{N} \mid x \geq 1$
Número de saídas(y)	$y \in \mathbb{N} \mid x \geq 1$
Número de camadas ocultas(h)	$h \in \mathbb{N} \mid x \geq 1$
Número de nós por camada oculta(n)	$n \in \mathbb{N} \mid x \geq 1$
Batch size(B)	$B \in \mathbb{N} \mid x \geq 1$
Épocas(E)	$E \in \mathbb{N} \mid x \geq 1$
Taxa de aprendizagem(η)	$\eta \in \mathbb{R} \mid 0 \leq \eta \leq 1$
Momentum(α)	$\alpha \in \mathbb{R} \mid 0 \leq \alpha \leq 1$

Tabela 2.3: Principais hiperparâmetros de uma MLP

também conhecido como *batch size*. Ele define a quantidade de amostras a serem usadas em uma iteração. O *batch size* será sempre um número inteiro, e a princípio quanto maior for esse número melhor (a depender da capacidade computacional). A tabela 2.3 resume os principais hiperparâmetros a serem considerados/ajustados para a modelagem de uma MLP.

O ajuste dos hiperparâmetros é a chave para a modelagem da rede. Entretanto, não há uma fórmula exata para preencher a coluna valor da tabela 2.3, exceto pelos valores de x e y como já foram citados previamente. Durante o treinamento da rede, o uso de métricas podem auxiliar na análise do resultado da rede e apontar a necessidade de se reajustarem alguns parâmetros. Esse processo é empírico, isto é, baseado na tentativa e erro, mas para efeito de atribuição de parâmetros iniciais, vamos dividi-los em grupos conforme Figura 2.32.

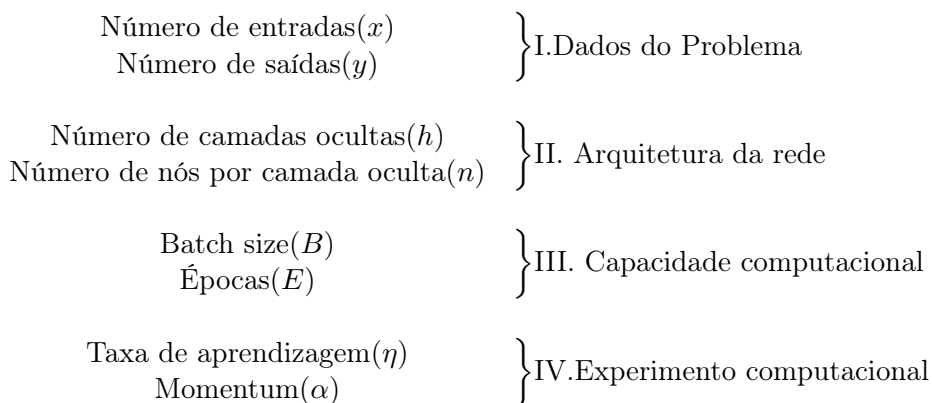


Figura 2.32: Hiperparâmetros em grupos
fonte:autor

O grupo I de hiperparâmetros poderá ser deduzido a partir dos dados do problema, isto é, o enunciado ou problema proposta irá evidenciar quais os possíveis resultados(saídas) bem como as entradas pelas características que definem o objeto de estudo. Por exemplo: uma rede de classificação de insetos poderia ter como saídas *formiga* e *aranha*, e como entradas *número de patas* e *comprimento*.

O grupo II está diretamente ligado ao projeto de arquitetura da rede. Quanto mais complexo o problema, a princípio mais camadas ocultas e nós ocultos são necessárias para que haja mais interações. Entretanto, interações demais podem afetar a complexidade dos cálculos da rede de maneira desnecessária. Portanto, o ideal é iniciar com uma quantidade menor de camadas e observar a curva de aprendizagem. A depender do problema, pode-se iniciar por uma arquitetura de um problema semelhante resolvido anteriormente.

O grupo III é formado por parâmetros inteiramente delimitados pela capacidade computacional. Quanto mais épocas e batch size melhor. Portanto, deve-se usar o maior valor possível para a capacidade computacional disponível.

A taxa de aprendizagem e o momentum (Grupo Iv) são parâmetros que devem ser monitorados e ajustados conforme a curva de aprendizagem vai se adequando. O experimento de Haykin dá uma boa ideia de como se atribuir valores iniciais para esses parâmetros.

2.8 A rede convolucional

A Rede Neural Convolucional (RNC) ou *Convolutional Neural Network* (CNN) é um tipo de RNA que, assim como a MLP, possui camadas ocultas, que neste caso são conhecidas como camadas convolucionais. As camadas convolucionais de uma CNN são especializadas em identificar um tipo de padrão, o que a torna especialmente eficaz para reconhecimento de imagens. De acordo com O'Shea & Nash (2015): "A única diferença notável entre uma CNN e uma RNA tradicional é que a CNN é usada primariamente para reconhecimento de padrões em imagens".

O que as camadas convolucionais fazem, apesar do nome, é uma operação de correlação-cruzada, também conhecida como produto escalar deslizante (sliding dot product). Uma aplicação conhecida como "*DEEPLIZARD Demo*" disponível no site <<https://deeplizard.com/resource/pavq7noze2>> pode ajudar a entender como ocorre a convolução (correlação-cruzada) em uma CNN.

A aplicação adota como padrão uma entrada 28x28 e uma saída de 26x26. No menu de opções da aplicação (Figura 2.34) foi escolhido o dataset 'MNIST' e o valor de entrada '0'. O valor padrão do filtro é 3x3 e o step é 1. Isso quer dizer que uma matriz 3x3 representará um filtro que vai deslizar de uma em uma célula. O dataset MNIST possui diversas imagens rotuladas de dígitos manuscritos e é uma base muito boa para treinar um modelo de reconhecimento de caracteres. Na aplicação estão disponíveis apenas dígitos 0 ao 9. Assim que a entrada '0' é escolhida o quadro de entrada 28x28 apresenta a imagem demonstrada na Figura 2.33.

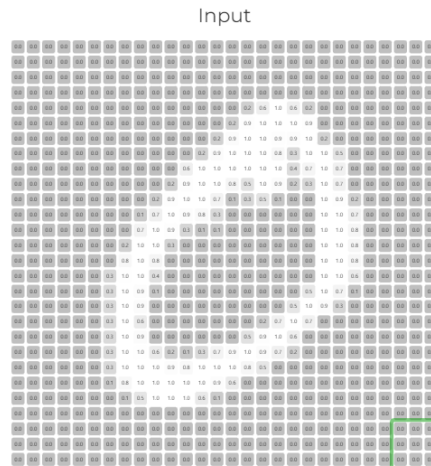


Figura 2.33: Entrada do valor 0 - MNIST
 fonte:([DEEPLIZARD... , 2021](#))

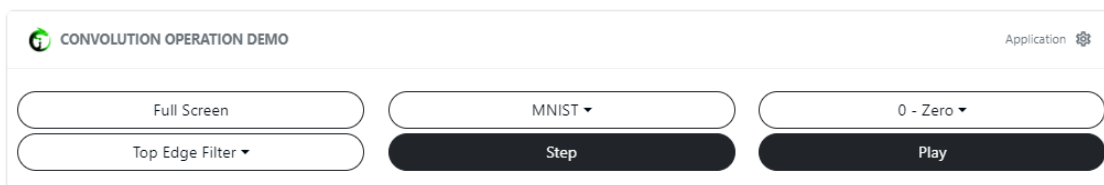


Figura 2.34: Menu DEEPLIZARD Demo
 fonte:([DEEPLIZARD... , 2021](#))

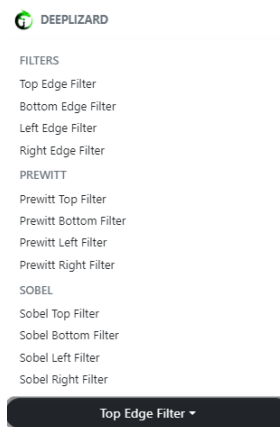


Figura 2.35: DEEPLIZARD Demo- filtros
 fonte:([DEEPLIZARD... , 2021](#))

A aplicação possui um total de doze filtros (Figura 2.35). Para esse exemplo, foram utilizados os seguintes: "Top Edge Filter", "Bottom Edge Filter", "Left Edge Filter" e "Right Edge Filter". Esses filtros irão detectar padrões de bordas no topo, na base, no lado esquerdo e lado direito respectivamente.

Abaixo do menu é possível se observar a operação realizada entre a janela deslizante e a entrada, bem como seu resultado na caixa output (Figura 2.36). Percebe-se que para o filtro selecionado "Top Edge Filter", a matriz filter possui os valores -1 na primeira



Figura 2.36: Produto Escalar - filtro
 fonte:(DEEPLIZARD. . . , 2021)

linha, +1 na segunda linha e zero na terceira. Isso fará com que, quando a janela passe, por exemplo, pela primeira borda de topo com valores 0.6, 1.0 e 0.6 a operação retome valores maiores nessa posições. No caso do filtro "Left Edge Filter", seguindo o mesmo raciocínio, é de se esperar que os valores -1 se concentrem na primeira coluna. Dessa forma, as equações 2.20, 2.21, 2.22 e 2.23 representam as matrizes dos quatro filtros selecionados.

$$\text{Top Edge Filter} \begin{bmatrix} -1 & -1 & -1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.20)$$

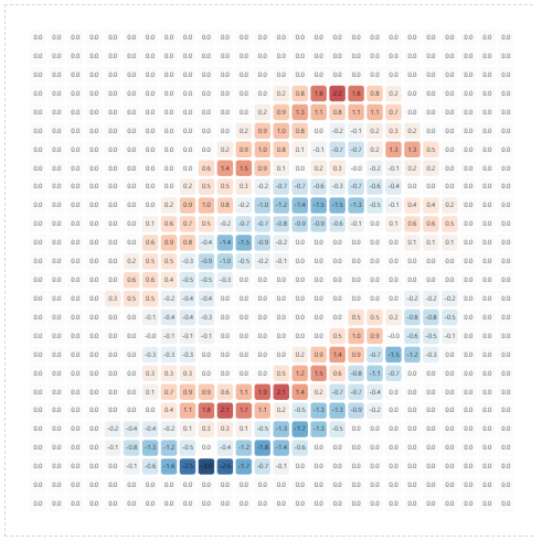
$$\text{Left Edge Filter} \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} \quad (2.21)$$

$$\text{Right Edge Filter} \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix} \quad (2.22)$$

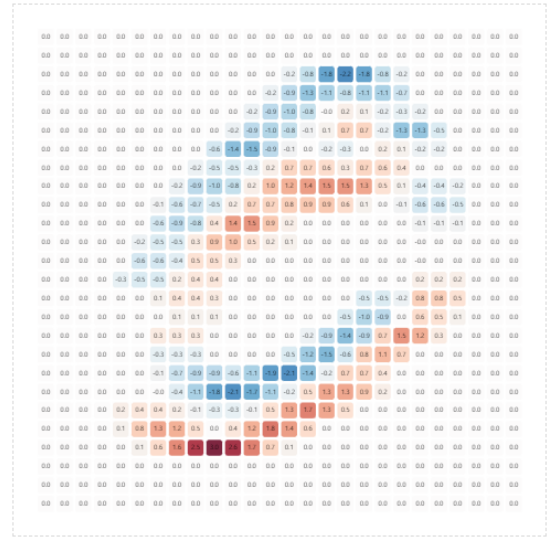
$$\text{Bottom Edge Filter} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \quad (2.23)$$

Após deslizar a janela por todo quadro input (Figura 2.33) com todos os quatro filtros, obtém-se as imagens dos quadros outputs da Figura 2.37. Os maiores valores obtidos da operação estão demarcados em vermelho, os menores valores demarcados em azul e os valores neutros permanecem em branco.

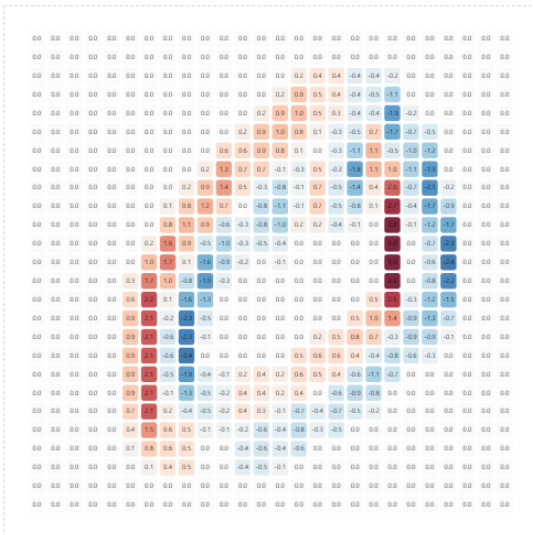
Com isso, é possível entender como que a operação da camadas convolucionais conseguem



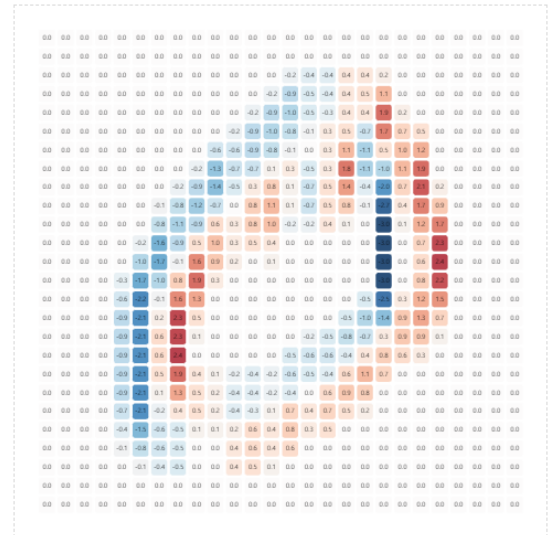
(a) Filtro Top Edge



(b) Filtro Bottom Edge



(c) Filtro Left Edge



(d) Filtro Right Edge

Figura 2.37: Saídas de filtros de convolução fonte:(DEEPLIZARD., 2021)

detectar padrões de borda. Se nas imagens das Figuras 2.37a,2.37b,2.37c e 2.37d forem deixados apenas os valores em vermelho, e posteriormente essas imagens forem sobrepostas, será possível se verificar uma reprodução muito fiel da entrada. É dessa forma que a CNN consegue realizar reconhecimento de padrões.

2.9 Métodos de detecção de múltiplos objetos

Uma determinada imagem capturada do ambiente e processada durante um mapeamento ou uma busca pode, muitas vezes, conter mais de um objeto de interesse, ou pode conter mais de um objeto de classe conhecida. Até o momento, os métodos apresentados para

reconhecimento de objetos baseados em aprendizagem de máquina tem sido reconhecidos por terem boas performances, entretanto se considerarmos uma imagem com um único objeto. Em 2014, um trabalho apresentado por Ross Girshick, Jeff Donahue, Trevor Darrell e Jitendra Malik intitulado "*Rich feature hierarchies for accurate object detection and semantic segmentation*" apresentou um novo método para detecção de objetos, capaz de resolver essa questão. O método foi chamado de "Region with CNN" ou, simplesmente, R-CNN. O sistema de detecção desenvolvido consiste de três módulos: o primeiro gera propostas de regiões, o segundo é uma CNN que extrai um vetor de características de comprimento fixo de cada região e o terceiro é um conjunto de SVMs² lineares de classe específica (GIRSHICK *et al.*, 2014).

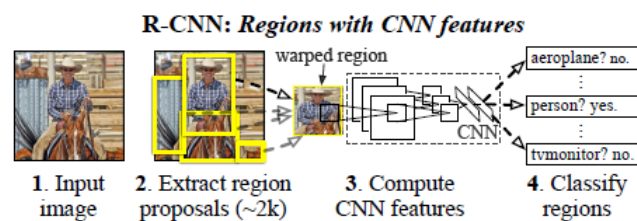


Figura 2.38: Detecção de objeto com R-CNN
 fonte:(GIRSHICK *et al.*, 2014)

A Figura 2.38 ilustra o processo do R-CNN. No primeiro quadro é possível se observar uma imagem com várias classes de objetos, por exemplo, cavalo, pessoa, etc. O segundo quadro mostra regiões propostas, das quais são extraídas propostas de região. Uma das regiões propostas contém o homem que será computado na rede CNN feature onde poderá ser identificado como pessoa. O mesmo poderá acontecer para o quadro que contém o cavalo caso tenha uma classe de interesse de animais.

Em 2015, Ross Girshick apresentou um novo método, que seria na verdade uma evolução do R-CNN, o qual chamou de Fast R-CNN. O Fast R-CNN seria uma formulação baseada no R-CNN onde com uma série de ajustes causariam uma melhoria nas velocidades de treinamento e de teste, bem como na acurácia da detecção (GIRSHICK, 2015).

Em 2016, Joseph Redmon, Santosh Divvala, Ross Girshick e Ali Farhadi apresentaram um novo método para detecção de objetos em um trabalho intitulado "*You Only Look Once: Unified, Real-Time Object Detection*". Esse novo método que ficou conhecido como YOLO (You Only Look Once) tem como ideia base unificar os componentes de detecção de objeto em uma única rede. Para isso, o sistema irá dividir uma imagem em uma grade $S \times S$ e para cada célula da grade irá prever B bounding boxes e C probabilidade de classe (REDMON *et al.*, 2016).

Na Figura 2.39 podemos ver que na imagem de entrada é feita segmentação da imagem

²Support Vector Machines(SVM), ou máquina de vetores de suporte é um método de aprendizagem supervisionada usada em problemas de classificação

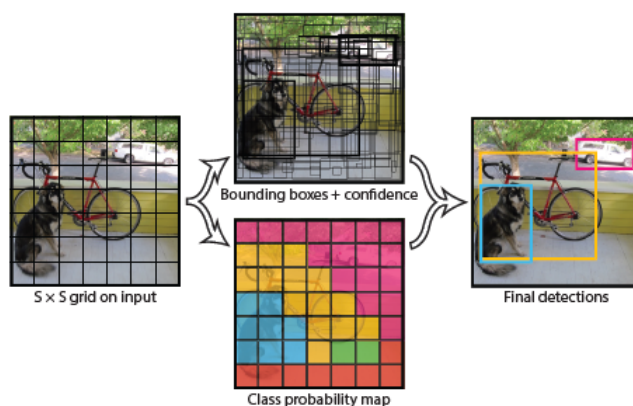


Figura 2.39: Detecção de objeto com YOLO
 fonte:(REDMON *et al.*, 2016)

em uma grade e que simultaneamente ocorrem as atribuições de bounding boxes com a probabilidade de classes. O resultado é que, conforme é possível ver no último quadro, poucas bounding boxes já com as classes designadas surgem. No exemplo da Figura 2.39, o sistema foi capaz de detectar rapidamente três objetos e entendê-los como cachorro, bicicleta e carro.

2.10 DL, ATR e a questão dos datasets

Nas seções anteriores, alguns dos principais métodos de aprendizagem de máquina e aprendizagem profunda foram apresentados. Recentemente, o uso de DL tem ganhado grande atenção no campo do reconhecimento de alvo. A acurácia do reconhecimento, nesse caso, pode ser melhorada através de um modelo de treinamento com um grande número de amostras. Entretanto, quando se trata de alvos abaixo d'água, sobretudo alvos perigosos que possuem formas diversas, as amostras para treinamento podem ser insuficientes para se obter um bom resultado(TENG; ZHAO, 2020).

De acordo com Jin & Liang (2017), o uso de Redes Neurais Profundas, ou seja, *Deep Neural Networks* (DNN), facilmente sofrem *overfitting* ao treinar modelos com pequenas amostras e no caso de amostras de imagens subaquáticas, é difícil se obter uma boa amostra de imagens rotuladas devido ao alto esforço e alto custo de aquisição envolvidos. Além disso, há as complexidades do ambiente marinho que dificultam a aquisição de imagens que foram citadas na seção 2.4.2. Essas questões levaram os pesquisadores a buscar várias maneiras alternativas de tratar o uso de datasets e os métodos de treinamento das redes, a fim de contornar essa limitação e tirar o proveito máximo do potencial das técnicas de DL.

Em face da insuficiência de *dataset* para o ambiente marinho, *transfer learning* (TL) pode ser usado para treinar um modelo que tenha uma grande quantidade de alvos em terra e no ar, para posteriormente transferir esse modelo para os alvos subaquáticos. (TENG; ZHAO, 2020). Yu *et al.* (2018) baseou-se na hipótese de que é possível adquirir imagens aéreas suficientes e fazer uso da combinação de DL e TL para desenvolver um novo sistema de reconhecimento de objetos com alto poder representativo e capaz de superar métodos tradicionais de *deep learning* para tarefa reconhecimento. Jin *et al.* (2019) propôs um modelo DCNN fim-a-fim para o desenvolvimento de ATR que nomeou de EchoNet. O modelo DCNN e rede de treinamento foram baseados em *transfer learning*, fazendo-se uso de um dataset grande.

Uma outra saída para a dificuldade de *dataset*, apresentada por Creswell *et al.* (2018), seria o uso de *Generative Adversarial Networks* (GANs), como uma forma de aprender representações sem treinar extensivamente dados rotulados. Os GANs conseguem isso derivando sinais de *backpropagation* através de um processo envolvendo um par de redes competitivas. As representações que podem ser aprendidas com GAN podem ser usadas em uma variedade de aplicações, incluindo síntese de imagem, edição de imagem semântica, *style transfer*, super resolução de imagem e classificação. O uso de GAN também foi destacado por Du *et al.* (2022), que fez uso desse recurso para prover um meio de *data augmentation* para amostras originais de sonar de abertura sintética, também conhecido como *synthetic Aperture Sonar* (SAS).

Huo *et al.* (2020a) fez uso de uma combinação de geração de dado semissintético e *deep transfer learning* como uma maneira de melhorar a acurácia da classificação de objeto abaixo d'água. Ele mostrou que ao transferir uma DCNN pré-treinada e fazer ajustes finos dados reais e semissintéticos para treinamento é possível obter resultados muito bons. Lee *et al.* (2019) apresentou um dataset gerado por simulação para treinamento de detecção de objetos abaixo d'água, com foco em geração de imagens SONAR sintéticas por transferência de estilo (*style transferred*).

2.11 Por que devemos olhar para o fundo do mar?

Nesta seção serão abordadas algumas das muitas situações que podem ser largamente favorecidas pelo desenvolvimento da tecnologia de reconhecimento de objetos abaixo d'água. Esta seção não esgota as possibilidades relacionadas ao método proposto.

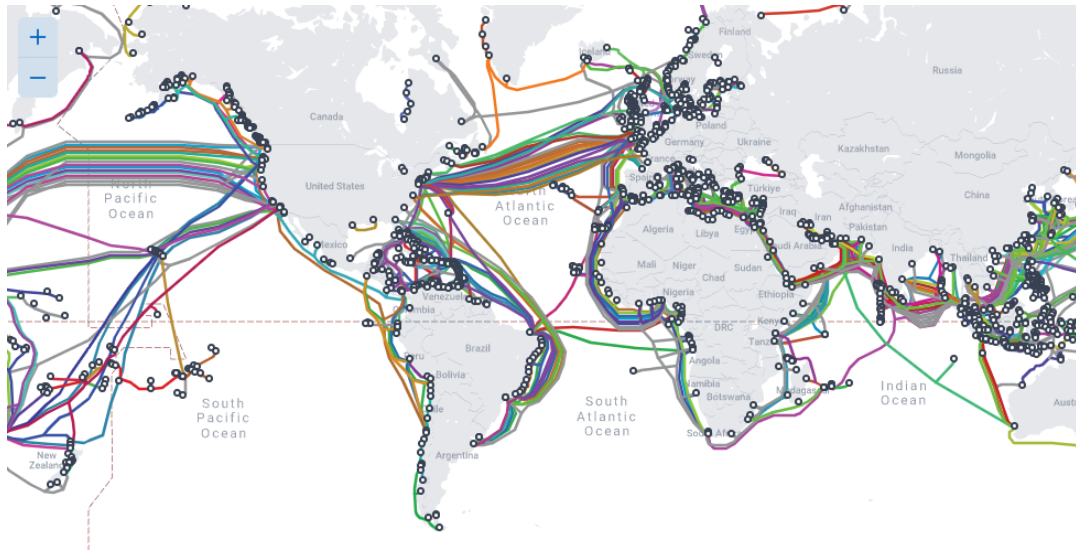


Figura 2.40: Cabos subamrinos no mundo
fonte: <<https://www.submarinecablemap.com/>>

2.11.1 Cabos submarinos: o mundo conectado

A Figura 2.40 foi obtida a partir do site <<https://www.submarinecablemap.com/>>. Nela, é possível se observar a distribuição dos cabos submarinos pelo mundo. De acordo com ANATEL (2022):

"Os cabos submarinos ancorados na costa brasileira atendem as necessidades do sistema de telecomunicações nacional e viabilizam a interconexão de qualquer sistema de telecomunicações e internet da América Latina com os demais continentes do mundo"

Ao observar a distribuição de cabos pela costa brasileira, é possível ver que existe um importante nó na cidade de Fortaleza-CE, do qual vários cabos da América do Sul convergem. A partir desse nó, um cabo de 6.200km da empresa Ellalink conecta o Brasil com Portugal, Marrocos e Ilha da Madeira, sendo esse nó, portanto, um ponto crítico. Outros importantes nós se encontram em Praia Grande-SP, Santos-SP e Rio de Janeiro-RJ. Um cabo de 11.000km identificado como "BRUSA" da empresa Telxius liga o Rio de Janeiro -RJ à Virginia Beach - VA nos Estados Unidos, estabelecendo conexões com Fortaleza-CE e San Juan na República Dominicana.

Os cabos e tubulações usados em uma grande variedade de cabos submarinos, principalmente nos nós que possuem grande convergência, fazem parte da infraestrutura crítica do país (ANATEL, 2022). Sendo assim, essa infraestrutura possui importância estratégica para o país.

Em fevereiro de 2020, um navio oceanográfico russo de alta tecnologia identificado como Yantar gerou alerta para a Marinha do Brasil (MB) por ter adentrado a Zona Econômica Exclusiva (ZEE) da costa brasileira. Essa mesma embarcação já havia gerado alertas em outros países e em 2018 foi escoltada pela Força Aérea britânica no canal da Mancha. De acordo com [G1 \(2020\)](#):

O que mais intrigou as autoridades náuticas foi o fato de a embarcação, que vinha do Uruguai, "reaparecer" perto dos cabos submarinos de comunicação que ligam o Brasil a outros países, após ficar por quase uma semana com o seu aparelho identificador desligado

Portanto, não só para efeito de manutenção e operação das comunicações, mas também, para fins estratégicos, é imprescindível que se haja um monitoramento, mapeamento e acompanhamento desses cabos que compõe a infra-estrutura crítica. O uso de sistemas automatizados de detecção e reconhecimento juntamente com embarcações remotas e/ou autônomas pode ser a solução para essa questão. [Eleftherakis & Vicen-Bueno \(2020\)](#) apresenta uma visão geral de tecnologias de sensoriamento já existentes que podem ser usadas com veículos autônomos para monitoramento de cabos subaquáticos.

2.11.2 Atividades econômicas

De acordo com a Convenção das Nações Unidas sobre o Direito do Mar (CNUDM), o Brasil possui os seguintes espaços marítimos:

- Mar Territorial: não deve ultrapassar o limite de 12 milhas náuticas (MN);
- Zona Contígua: adjacente ao mar territorial, cujo limite máximo é de 24 MN e é medida a partir das linhas de base do mar territorial;
- Zona Econômica Exclusiva (ZEE): medida a partir das linhas de base do mar territorial e que não deve exceder a distância de 200 MN;
- Plataforma Continental: A plataforma continental de um Estado costeiro compreende o leito e o subsolo das áreas submarinas que se estendem além do seu mar territorial, em toda a extensão do prolongamento natural do seu território terrestre, até ao bordo exterior da margem continental, ou até uma distância de 200 milhas marítimas das linhas de base a partir das quais se mede a largura do mar territorial, nos casos em que o bordo exterior da margem continental não atinja essa distância.

Na ZEE, o Brasil possui prioridade na utilização de recursos vivos e não-vivos. Conforme é possível observar na [Figura 2.41](#), a área em azul claro que corresponde à ZEE contém uma

grande quantidade de recursos ao longo de toda costa brasileira, a saber: areia, cascalho, diamante, enxofre, cobalto, níquel, cobre, magnésio, carvão, fósforo entre outros. Além de todos esses recursos, conta também com petróleo e gás natural que são importantes fontes de energia, assim como os recursos vivos, onde muitos deles são adequados para a atividade pesqueira.

Há diversas maneiras pelas quais as atividades econômicas realizadas no mar podem se favorecer de um sistema ATR: prospecção e localização de recursos subaquáticos, detecção e classificação de objetos vivos e não-vivos, mapeamento de distribuição de recursos ao longo da costa brasileira, apoio a atividades de detecção, avaliação e reparo de estruturas subaquáticas (no caso de petróleo e gás) dentre outros.

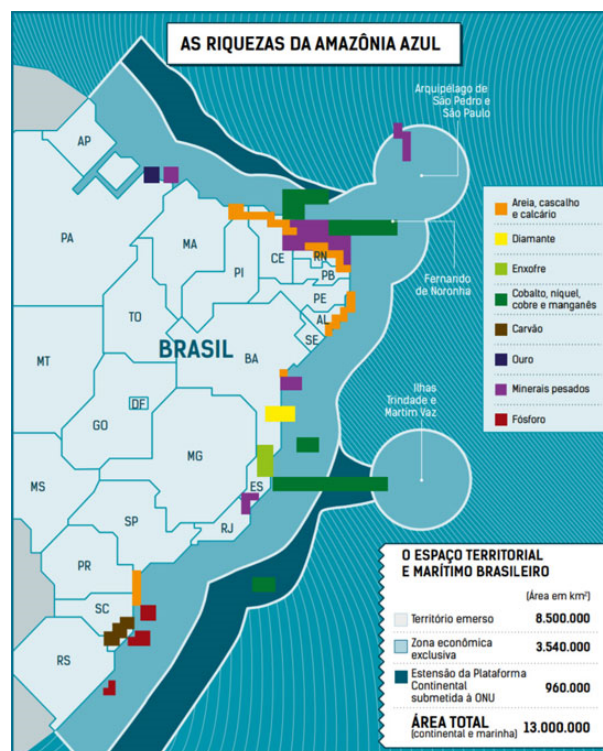


Figura 2.41: Recursos e atividades econômicas na Amazônia Azul
fonte:(PADILHA, 2020)

2.11.3 Arqueologia submarina

No fundo do mar mediterrâneo, em algum lugar a cerca de 6,5 km de Alexandria, jaz a antiga cidade egípcia de Thonis-Heracleion. Evidências apontam que ela fora um importante centro comercial, tendo sido uma cidade portuária. De acordo com Malva (2020):

"Também conhecida como Heracleion, a cidade fora construída em torno de seu grande templo, e era atravessada e interligada por uma rede de canais. Cada pequena ilha, entre as teias aquática, abrigava casas e pequenos santuários."



Figura 2.42: Achado arqueológico em Thonis-Heracleion
fonte:(MALVA, 2020)

A descoberta de Heracleion trouxe consigo uma grande diversidade de objetos arqueológicos, desde grandes estátuas (Figura 2.42), objetos de santuários, e até mesmo lápides com inscrições. Segundo G1 (2022), são exemplos de cidades submersas espalhadas pelo mundo: Baia na Itália, Thonis-Heracleion no Egito, Derwent na Inglaterra, Villa Epecuén, na Argentina e Port Royal, na Jamaica.

Em Pernambuco, a cerca de 400 km de Recife, a cidade de Petrolândia é conhecida como uma cidade parcialmente submersa, e conhecida pela população local como "Atlântida brasileira". A cidade antes conhecida como Jatobá, passou a se chamar Petrolândia em homenagem ao Imperador D. Pedro II que a visitou no ano de 1877. Após uma inundação ocorrida na década de 80 devido à construção da Usina Hidrelétrica Luiz Gonzaga, o centro da antiga cidade e vários povoados situados em áreas próximas foram alagados e a população foi reassentada em bairros da Nova Petrolândia, em agrovilas e projetos de irrigação (PRETOLÂNDIA-PE, 2017). Entretanto, em 2014, uma construção importante da velha Petrolândia, a igreja do Sagrado Coração voltou a aparecer (Figura 2.43), ficando parcialmente submersa e desde então, passou a se tornar um importante ponto turístico na região.

Achados arqueológicos, monumentos, antigas construções e templos submersos são objetos artificiais, que se diferenciam facilmente de objetos naturais por normalmente conterem formas geométricas bem definidas. Dessa forma, a busca por esse tipo de objeto, parte integrante importante da arqueologia, pode se favorecer da tecnologia de detecção e classificação. Moroni *et al.* (2014) apresentou um método de processamento de dados acústicos e ópticos capaz de identificar e realçar esses tipos de objetos no meio ambiente submarino.

Figura 2.43: Igreja submersa Sagrado Coração - Petrolândia



fonte: <gl.globo.com>

2.11.4 Busca e Salvamento no mar

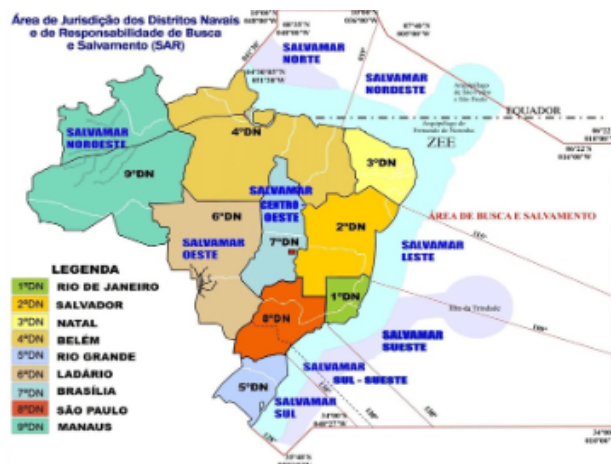


Figura 2.44: Áreas SAR- Marinha do Brasil

fonte: <<https://www.marinha.mil.br/dgn/?q=salvamar>>

O item 2 do artigo 98 da Convenção das Nações Unidas sobre o Direito do Mar (CNUDM) estabelece o seguinte:

Todo Estado costeiro deve promover o estabelecimento, o funcionamento, e a manutenção de um adequado e eficaz serviço de busca e salvamento para garantir a segurança marítima e aérea, e, quando as circunstâncias o exigirem, cooperar para esse fim com os Estados vizinhos por meio de ajustes regionais de cooperação mútua.

Em consequência do estabelecido na CNUDM, e por meio do Aviso Ministerial nº N-0201, de 20 de fevereiro de 1970, foi criado o Serviço de Busca e Salvamento da Marinha do Brasil (SALVAMAR). O SALVAMAR tem como objetivo a salvaguarda da vida humana

no mar, nos portos e nas vias navegáveis interiores. A área de responsabilidade da Marinha do Brasil para esses serviços é de cerca de 14,2 milhões de km² de área marítima, correspondentes a 1,6 vezes a área do território nacional, somados às vias navegáveis interiores.

Atualmente, as buscas no mar são realizadas pela Marinha do Brasil fazendo-se uso de meios distritais como navios-patrolhas e aeronaves de serviço. O tempo entre o acionamento do chamado de SAR (*Search and Rescue*) e o início da operação de busca depende do tempo de guarnecimento do navio de serviço e a distância que ele se encontra da localização da operação. Para atender toda a extensa área sob sua responsabilidade, a MB dividiu toda a área SAR conforme a Figura 2.44.

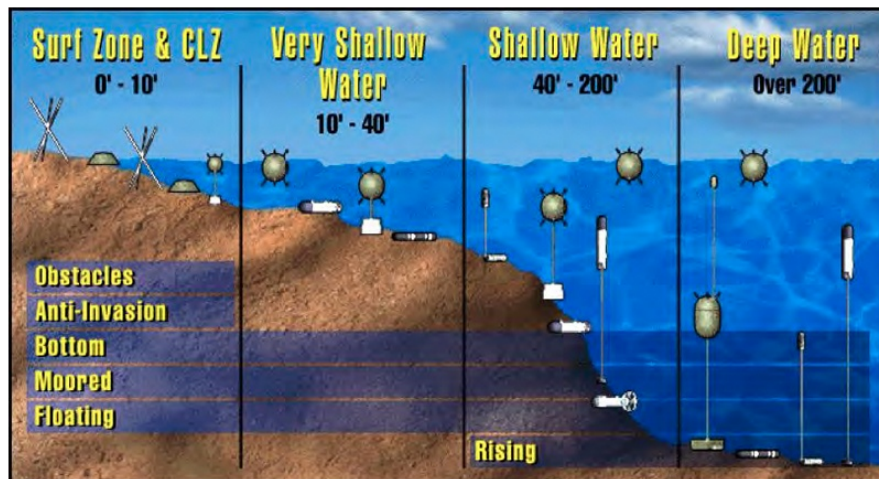
Uma vez que a CNUDM foi assinada por mais de 160 países, o problema do SAR é uma questão global e, portanto, em toda parte do mundo há muitos estudos a seu respeito. [Murphy et al. \(2008\)](#) apresenta um conceito para cenários de SAR usando veículos autônomos, a partir de tecnologias atualmente disponíveis.

Um projeto que ficou conhecido como "euRathlon 2015 Grand Challenge" foi realizado em Piombino-Itália entre 2013-2015 com o intuito de organizar competições de robôs em cooperação em cenários realistas de SAR e testar a inteligência e autonomia dos mesmos. [Matos et al. \(2016\)](#) apresentou os resultados da participação do INESC TEC (Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência) na euRathlon 2015 no ambiente marinho (superfície e abaixo d'água). Os robôs 'ROAZ USV' e 'MARES AUV' participaram de tarefas múltiplas como avaliação da situação, mapeamento abaixo d'água, detecção de vazamento e localização de vítimas.

As tecnologias em desenvolvimento deverão, em breve, redimensionar as operações de SAR. O uso de diversos robôs com capacidades distintas trabalhando em conjunto com os métodos convencionais deverão provocar um grande impacto na performance dessas operações. Mas, para isso, é fundamental o aprimoramento dos algoritmos de detecção, classificação e reconhecimento de objeto.

2.11.5 Detecção de minas submarinas

As minas submarinas são artefatos explosivos que podem ser usados ou na superfície ou em baixo d'água. Esses artefatos possuem características muito variadas podendo ser desde um arcaico artefato que libera carga explosiva com a pressão, como também dispositivos inteligentes escondidos a uma profundidade adequada sendo acionadas por influência de ondas acústicas ou magnéticas (Figura 2.45). Elas podem ser usadas tanto defensivamente, protegendo o mar territorial de ataques externos, como também



Source: Taking Mines Seriously, Scott C Truver.

Figura 2.45: Tipo de minas subaquáticas
 fonte:([TRUVER, 2012](#))

ofensivamente, ou seja, no mar territorial inimigo.

Por ser um recurso de custo relativamente baixo capaz de anular/neutralizar as ações dos mais modernos navios de guerra, as minas submarinas são um elemento importante para uma marinha de guerra, gerando assim a necessidade de se desenvolver tanto as atividades de minagem como também as suas contra-medidas, isto é, as contra-medidas de minagem. De acordo com [Naval \(2003\)](#):

"As contramedidas de minagem (CMM) têm por finalidade atenuar ou, se possível, anular esses esforços desenvolvidos pelo inimigo. As CMM compreendem todos as técnicas e métodos empregados para se oporem à ameaça representada pelas minas já lançadas pelo inimigo."

Diferentemente dos artefatos explosivos, as operações de CMM são complexas, dispendiosas, e envolvem muito risco de pessoal e material. Entretanto, há uma tendência atual de se modernizar as doutrinas de CMM, buscando reduzir o pessoal envolvido nos riscos dessas operações. Para isso, as CMM podem ser realizadas por veículos não-tripulados. Nesses casos, esses veículos podem realizar missões de forma autônoma ou remotamente, equipados com sensores de alta resolução e os dados coletados são usados para detecção, classificação e identificação ([HOŽYŇ, 2021](#)). [Khaledi et al. \(2014\)](#) realizou simulações de dois sonares e cinco veículos autônomos distintos para avaliar a performance desses sistemas na detecção de minas subaquáticas. Ele utilizou um modelo computacional para simular a dinâmica e o consumo de combustível de cada veículo.

Em geral, há dois tipos de navios que realizam CMM : navio-varredor e navio caça-minas.

O navio varredor é capaz de 'limpar' uma área em um tempo razoável, usando mais de um método simultaneamente, e estabelecendo um canal varrido, isto é, uma passagem segura para outros navios. Operações com estes tipos de navios são complexas e envolvem muito risco de pessoal. Os navios caça-minas, por sua vez, já possuem ferramentas semelhantes a ROV, podendo reduzir o risco de pessoal, porém suas operações são mais cirúrgicas atuando em áreas mais específicas e pra reduzir o risco de pessoal deverá dispor de mais tempo para estabelecer um canal varrido. Portanto, a fim de se desenvolver métodos mais eficientes e , ao mesmo tempo, a redução de risco pessoal, muitas marinhas estão desenvolvendo uma nova doutrina "Unmanned"(não-tripulada) e apostando no uso de SUV, AUV, e em alguns casos até mesmo gliders para neutralizarem ameaças desse tipo. Entretanto, é fundamental que esses veículos possuam a capacidade de detecção dos artefatos explosivos abaixo d'água.

A ideia de um algoritmo de grande acurácia, automatizável e de fácil implementação em um veículo que possa operar remotamente, ou de forma autônoma, constitui a chave-mestra para o desenvolvimento das tecnologias da caça-de-minas que vêm se tornando mais requisitadas a cada dia, conforme as minas vão se tornando cada vez mais inteligentes. De acordo com [Teng & Zhao \(2020\)](#): "A acurácia do reconhecimento de alvos abaixo d'água por AUV é uma poderosa garantia para detecção, resgate e segurança abaixo da água".

A detecção e classificação de objetos abaixo d'água é uma questão central no desenvolvimento das técnicas de caça-minas. Para desenvolver essas capacidades, é necessário não apenas ter em mãos um bom conjunto de sensores capazes de coletar grandes quantidades de dados com qualidade, mas também, ser capaz de tratar e utilizar esses dados de forma eficiente. Isto requer o uso de um algoritmo que possa tirar o maior proveito possível dos dados coletados e que seja, preferencialmente, automatizável.

Dessa forma, o desenvolvimento de um modelo computacional confiável pode ser considerado um passo fundamental no aperfeiçoamento das CMM, em consequência do incremento das capacidades de caça-minas, especialmente, considerando a adequação às novas tendências e doutrinas, bem como o surgimento de algoritmos de aprendizagem de máquina capazes de otimizar os algoritmos de detecção e classificação.

Materiais e métodos

Este capítulo apresenta o desenvolvimento do modelo computacional para reconhecimento de objetos em ambiente subaquático baseado em aprendizagem de máquina.

3.1 Dataset

O primeiro passo no desenvolvimento do modelo computacional foi a pesquisa por bases de dados de imagens subaquáticas existentes, a fim de se montar um *dataset* para treinamento e validação do modelo. A partir do *dataset*, é feita a escolha de qual arquitetura de RNA seria mais adequada para o criação do modelo. Nesse sentido, uma pesquisa inicial foi realizada com o auxílio do recurso do *google* intitulado "*Dataset Search*", disponível em: [<https://datasetsearch.research.google.com/>](https://datasetsearch.research.google.com/). As *tags* utilizadas foram: "*Underwater images*", "*underwater objects*", "*underwater detection*".

Para cada uma das *tags* utilizadas, foram encontrados mais de 100 resultados. Ao fim dessa pesquisa, considerando os critérios qualidade das imagens, disponibilidade do dataset para ser baixado e variedade de objetos (considerando tanto objetos naturais, quanto artificiais), foram pré-selecionados três *datasets*:

1. O *dataset* SeabedObjects-KLSG (Figura 3.1) é o resultado de um estudo com foco em classificação de vítimas de afogamento, naufrágios, aeronaves, minas e solo marinho. Atualmente contém 385 imagens de naufragio, 36 de vítimas de afogamento, 62 de aeronave, 129 de minas e 578 de solo marinho. As imagens foram obtidas a partir do emprego de sonar sidescan (HUO *et al.*, 2020b).
2. O *dataset* MARIS (Figura 3.2) foi construído a partir do projeto MARIS (*Marine Autonomous Robotics for InterventionS*) originário do RIMLAB (*Robotics and Intelligent Machines Laboratory*) da *Università di Parma*. O objetivo desse projeto é o desenvolvimento de AUVs para intervenções de operações no fundo do mar voltadas para a indústria *offshore*, busca e salvamento e explorações científicas diversas. O *dataset* foi obtido a partir de capturas de imagens a 15 frames por segundo, e contém 9600 imagens estéreo contendo sequência de imagens de tubos com diâmetro variando entre 5cm e 6cm (OLEARI *et al.*, 2015; KALLASI *et al.*, 2015; RIZZINI *et al.*, 2015).
3. O *dataset* UOT100 (*Underwater Object Tracking*) (Figura 3.3) foi criado para auxiliar no desenvolvimento de algoritmos de acompanhamento em ambientes subaquáticos.

O *dataset* consiste em 104 seqüências de videos e mais de 74000 *frames* de imagens naturais e artificiais (PANETTA *et al.*, 2022; KEZEBOU *et al.*, 2019).

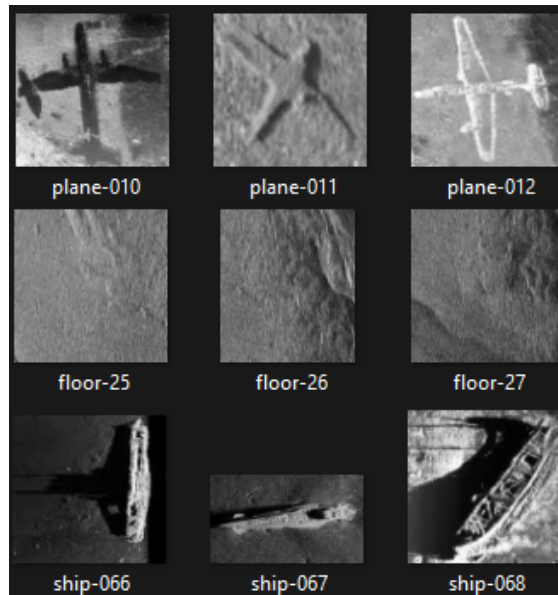


Figura 3.1: Amostra de imagens do dataset SEABED-KLSG-II
fonte: SeaBed-KLSG-II

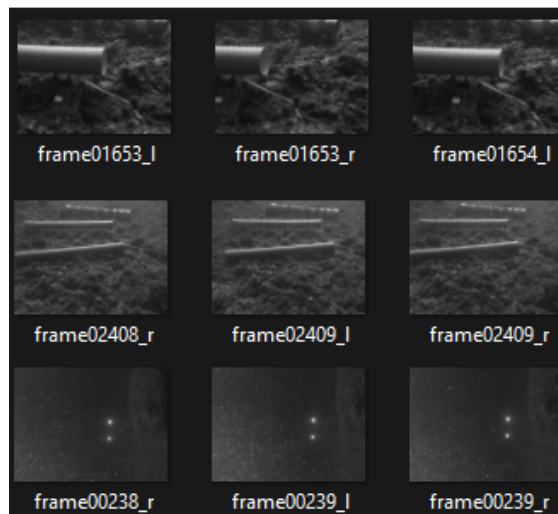


Figura 3.2: Amostra de imagens do dataset MARIS-Project-Set1
fonte: MARIS-Project

3.2 Modelo computacional de detecção e classificação

A Figura 3.4 ilustra o processo de criação do modelo computacional

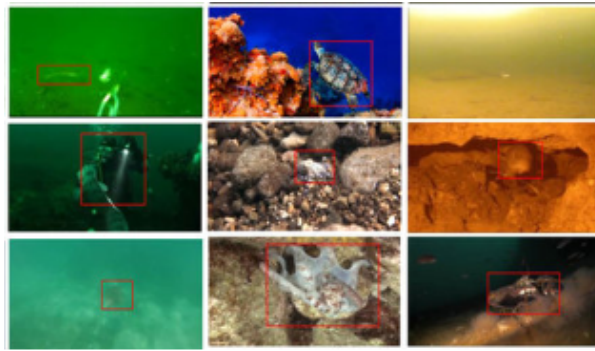


Figura 3.3: Amostra de imagens do dataset UOT-100
fonte: UOT-100

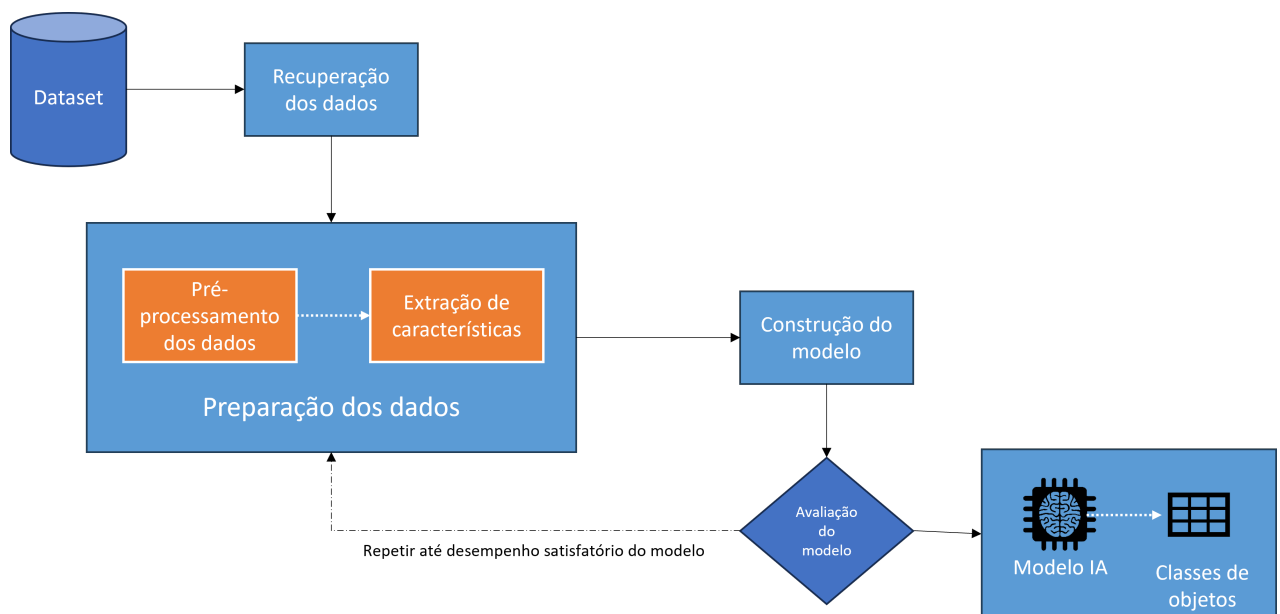


Figura 3.4: Etapas envolvidas na criação de modelo computacional.
Fonte: Autor.

O desenvolvimento do modelo ocorrerá através das seguintes etapas:

1. *Dataset*: para a criação e validação do modelo, foram selecionadas três categorias de imagens: tubo (pois se aproxima de objetos comumente encontrados durante inspeções marinha) e duas classes de animais marinhos tartaruga e tubarão (para que o modelo possa reconhecer padrões desses tipos de objetos). Nesse sentido, 300 imagens de cada classe foram selecionadas de forma aleatória para composição do *dataset*.
2. Recuperação de dados: consiste na leitura das imagens brutas onde é realizada a conversão para uma matriz bidimensional contendo os valores dos *pixels* da imagem.
3. Preparação dos dados: Esta etapa contém três subetapas, que são pré-processamento

dos dados, extração de característica e seleção de recursos. Para a realização do processamento dos dados foi utilizada a biblioteca ImageJ, trata-se de um *software* de código aberto para processamento e análise de imagens científicas (SCHINDELIN *et al.*, 2015).

- Pré-processamento: esta etapa é responsável pela aplicação do filtro de *Sobel* nos dados da matriz que representa a imagem, gerando uma nova matriz binarizada.
 - Extração de característica: o processo de extração de características, calcula os atributos que contém os padrões de entrada para o vetor de características. O vetor de características é composto por quatro atributos, que contém o padrão a ser analisado pelo processo de avaliação do modelo. Esses atributos são definidos como: *areaFraction* (porcentagem de *pixels* diferentes de zero), *mean* (média aritmética), *stdDev* (desvio padrão das intensidades de *pixel*) e área (área em *pixels* quadrados).
4. Construção do modelo: o modelo foi construído utilizando uma RNA do tipo *Multi Layer Perceptron* (MLP), com o algoritmo de retropropagação do erro (*backpropagation*). Nessa etapa, o modelo é treinado a partir de conjunto pré-determinado de classes: *pipe*, *turtle* e *shark*. Dessa forma, o sistema extrai os atributos do vetor de características para realizar o treinamento do classificador, através do conjunto de dados de treinamento. Aos elementos de treinamento, são associados rótulos de classes às quais cada um pertence. Esta etapa é conhecida como aprendizagem supervisionada, pois o rótulo da classe de cada elemento de treinamento é fornecido ao classificador. A saída do classificador indicará um comando para uma aplicação externa. Para a criação da RNA foi utilizada a biblioteca Weka, trata-se de um *software* de código aberto para criação de algoritmos de aprendizado de máquina (EIBE *et al.*, 2016).
 5. Avaliação do modelo: esta é a etapa onde faz-se o ajuste de hiperparâmetros e também, se o desempenho do modelo não for satisfatório, reitera o processo de Preparação dos dados.

3.3 Método de validação do modelo

Para realizar o treinamento do modelo computacional baseado numa RNA/MLP, o *dataset* foi dividido em 10 blocos para validação cruzada. Este processo consiste em particionar o conjunto de dados em um determinado número n de seções, com aproximadamente a mesma quantidade de dados, de modo a realizar repetidos treinamentos utilizando $n - 1$ conjuntos, para então realizar a validação dos dados com o subconjunto restante (PAL; PATEL, 2020).

Este processo repete-se n vezes, para que a cada treinamento uma nova combinação de subconjuntos seja utilizada para treinar a rede e, conseqüentemente, validar a rede com outros dados.

O resultado do modelo computacional é dado pela média entre os valores obtidos por todas as validações, o que contribui para uma melhor acurácia do modelo, uma vez que múltiplas combinações entre os dados foram feitas tanto para treinamento como para validação.

Resultados e discussão

Neste capítulo, será apresentado o experimento computacional realizado para avaliar o comportamento do modelo, relacionado ao desempenho da RNA.

O computador utilizado para a execução do modelo computacional possui um processador Intel Core i7 6700K com 4 núcleos e 8 *threads*, limitado a uma frequência de 4.20 GHz, 16 GB de RAM, e o sistema operacional Linux baseada no Ubuntu 20.04 LTS. Foi instalado o Weka na versão 3.8.6.

A RNA utilizada é composta por cinco camadas (entrada, duas ocultas e saída). O número de neurônios da camada de entrada da RNA é determinado pela dimensionalidade do vetor de características, sendo quatro. A quantidade de neurônios na camada de saída da rede foi definida como três neurônios, visto que o objetivo da rede é discriminar entre três objetos.

Segundo [Heaton \(2008\)](#), duas camadas escondidas já são capazes de representar qualquer relação entre os dados, mesmo aquelas que não podem ser representadas por equações. Mais do que duas camadas escondidas só são necessárias em problemas ainda mais complexos como séries temporais e visão computacional. Nesse sentido, a RNA foi modelada com duas camadas ocultas. Já o número de neurônios em cada camada é uma questão mais empírica, não existindo regras explícitas para um cálculo ideal. Nesse sentido, foi utilizado a metodologia proposta por [Han & Kamber \(2011\)](#), definida pela seguinte equação 4.1:

$$N_{hidden} = 2N_{in} + 1 \quad (4.1)$$

Onde N_{in} é o número de neurônios da camada de entrada e N_{hidden} representa o número de neurônios da camada oculta. Nesse sentido, o número de neurônios da camada oculta foi definido com nove. A arquitetura final da RNA pode ser visualizada na Figura 4.1.

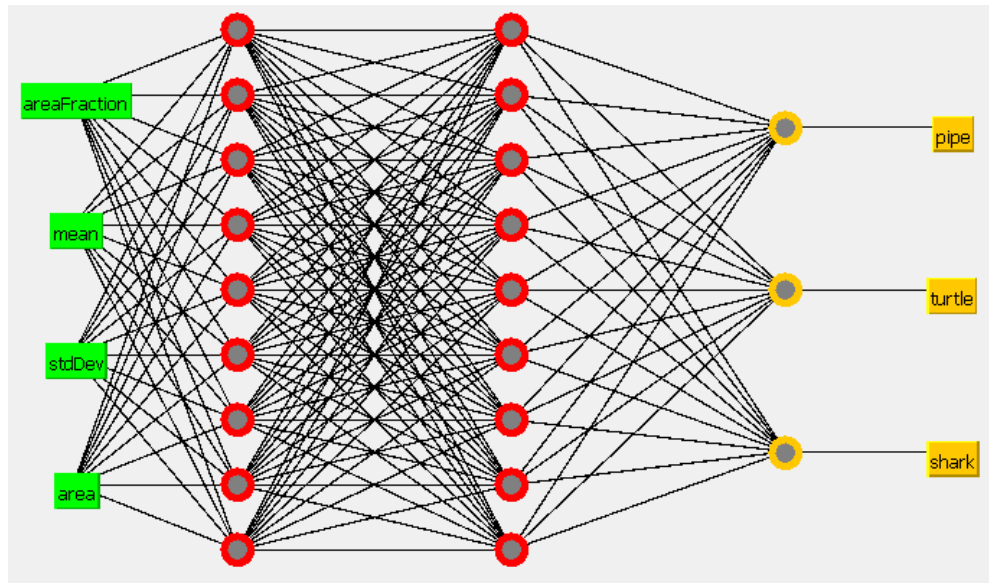


Figura 4.1: Arquitetura da rede neural utilizada.

Fonte: Autor.

Uma vez construída a RNA, sua precisão e capacidade de generalização foram testadas utilizando os dados processados das imagens subaquáticas que compõem o conjunto de amostras para treinamento e testes. Foram simuladas combinações do parâmetro da taxa de aprendizagem (0.01, 0.1, 0.5, 0.9) e da constante de *momentum* (0.0, 0.1, 0.5, 0.9) para observar o efeito da convergência da RNA. Cada combinação foi treinada até 1000 épocas.

As curvas de aprendizagem médias são ilustradas nas Figuras 4.2, 4.3, 4.4 e 4.5. Agrupadas pela taxa de aprendizagem (η) em função das constantes de *momentum* (α).

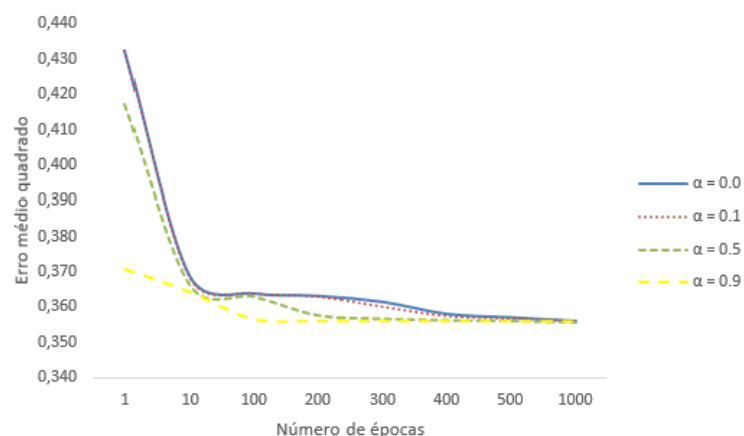


Figura 4.2: Curvas de aprendizagem médias para $\eta = 0.01$.

Fonte: Autor.

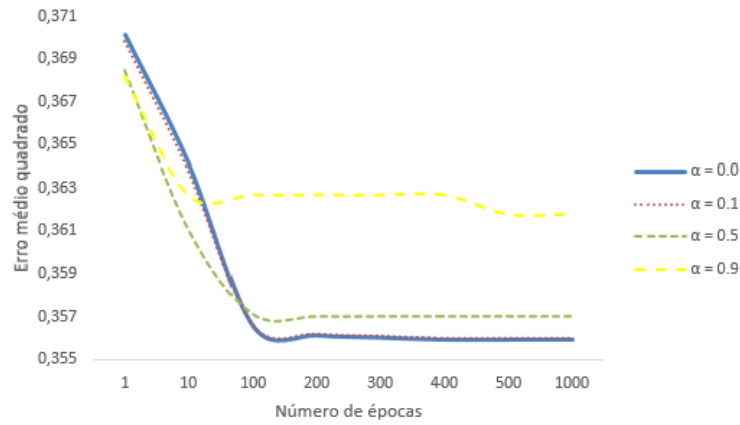


Figura 4.3: Curvas de aprendizagem médias para $\eta = 0.1$.
Fonte: Autor.

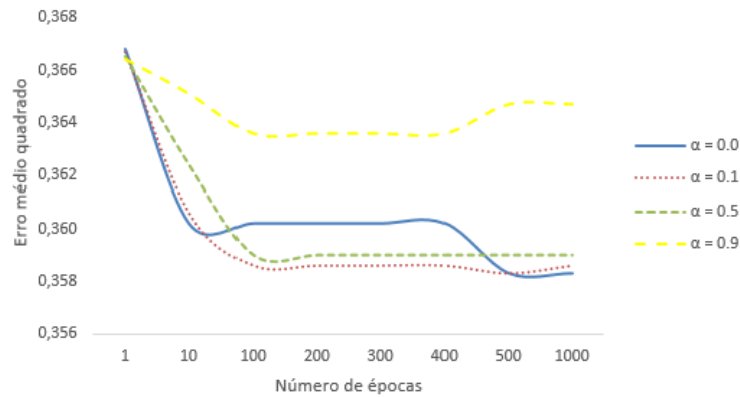


Figura 4.4: Curvas de aprendizagem médias para $\eta = 0.5$.
Fonte: Autor.

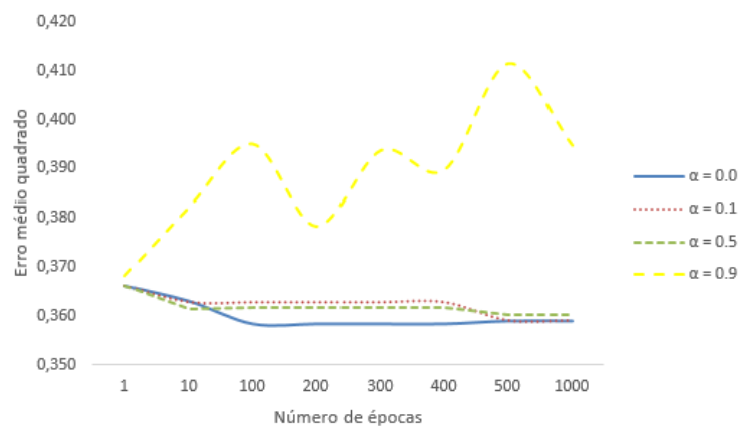


Figura 4.5: Curvas de aprendizagem médias para $\eta = 0.9$.
Fonte: Autor.

Observa-se, a partir dos resultados, ilustrados nas Figuras 4.2, 4.3, 4.4 e 4.5, que em geral, para uma taxa de aprendizagem (η) = 0.01, obtém-se uma convergência mais lenta. Para as taxas de aprendizagem $\eta = 0.5$ e $\eta = 0.9$, sendo $\alpha = 0.9$, observa-se oscilações no erro

médio quadrado durante o processo de aprendizagem, ocasionando um valor mais alto do erro médio quadrado na convergência da RNA, sendo ambos efeitos indesejáveis.

A Figura 4.6 apresenta as melhores curvas de aprendizagem para cada grupo das curvas ilustradas nas Figuras 4.2, 4.3, 4.4 e 4.5, para determinar a melhor curva de aprendizagem global.

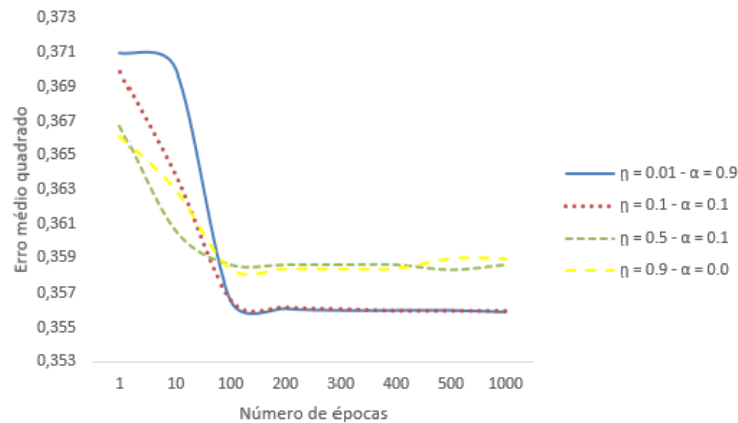


Figura 4.6: Melhores curvas de aprendizagem selecionadas.

Fonte: Autor.

Observa-se, a partir da Figura 4.6, que o melhor parâmetro da taxa de aprendizado foi $\eta = 0.1$ e a constante de *momentum* foi $\alpha = 0.1$. Nesse contexto, o fato de o erro médio quadrado da curva não variar muito, sugere que os parâmetros são os mais adequados para esse problema.

Nesse contexto, a RNA final foi avaliada para determinar a sua precisão e capacidade de generalização. Findado o experimento, a RNA demonstrou um percentual de instâncias classificadas corretamente igual a 98,7%. O número de instâncias classificadas corretamente foi de 888, contra 12 instâncias classificadas incorretamente. Ao analisarmos a matriz de confusão, na Tabela 4.1, observa-se, que a classe *pipe* todas as instâncias classificadas corretamente. As classes *turtle* e *shark*, obtiveram 294 instâncias classificadas corretamente, contra 6 instâncias classificadas incorretamente.

Tabela 4.1: Matriz de confusão da classificação dos dados.

Classes	<i>pipe</i>	<i>turtle</i>	<i>shark</i>
<i>pipe</i>	300	0	0
<i>turtle</i>	0	294	6
<i>shark</i>	0	6	294

Após a conclusão dos experimentos, de acordo com os resultados, observa-se que a RNA apresentou precisão média de 98,7% no processo de reconhecimento de padrões a partir

dos dados do *dataset*.

Considerações Finais

5.1 Conclusões

O modelo proposto foi capaz de identificar três tipos de objetos subaquáticos. Para obter este resultado, uma etapa inicial de processamento foi necessária para transformar os dados, extrair características estatísticas relevantes dos dados fornecidos para treinar e testar o modelo.

5.2 Contribuições

As principais contribuições desta pesquisa são:

- Desenvolvimento de algoritmos de pré-processamento, extração de característica e classificação para detecção de três objetos subaquáticos.
- Avaliação de desempenho e acurácia do modelo de modo a determinar a arquitetura mais eficiente para a solução do problema.

5.3 Atividades Futuras de Pesquisa

A utilização de *Deep Learning* está definida como atividade futura desta pesquisa. Estes algoritmos são ideais para reconhecimento de padrões e podem ser aplicados para o problema descrito neste trabalho. Deve-se avaliar a acurácia desta proposta, assim como o custo computacional para o treinamento e validação do modelo.

Outra proposta de atividade futura, seria a introdução de dados de novos objetos subaquáticos para ampliar a possibilidade de aplicações desse modelo computacional.

Referências Bibliográficas

- ANATEL. *Cabos submarinos*. 2022. <<https://www.gov.br/anatel/pt-br/dados/infraestrutura/cabos-submarinos>>. Accessed: 2023-03-15. 2.11.1
- BARELLI, Felipe. *Introdução à visão computacional: Uma abordagem prática com Python e OpenCV*. [S.l.]: Editora Casa do Código, 2018. 2.3, 2.7, 2.3
- COMMERCE, U.S. Department of. *NOOA Ocean Exploration*. 2001. Url=<https://oceanexplorer.noaa.gov/about/welcome.html>. Accessed: 2023-02-15. 2.8, 2.9, 2.10
- CRESWELL, Antonia *et al.* Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, v. 35, n. 1, p. 53–65, 2018. 2.10
- CUI, Weicheng. An overview of submersible research and development in china. *Journal of Marine Science and Application*, Springer, v. 17, n. 4, p. 459–470, 2018. 2.4.1
- DEEPLIZARD Demo. 2021. <<https://deeplizard.com/resource/pavq7noze2>>. Accessed: 2023-02-15. (document), 2.33, 2.34, 2.35, 2.36, 2.37
- DU, Shaoyan *et al.* A high-quality multicategory sar images generation method with multiconstraint gan for atr. *IEEE Geoscience and Remote Sensing Letters*, v. 19, p. 1–5, 2022. 2.10
- EIBE, Frank; HALL, Mark A; WITTEN, Ian H. The weka workbench. online appendix for data mining: practical machine learning tools and techniques. In: *Morgan Kaufmann*. [S.l.: s.n.], 2016. 4
- ELEFTHERAKIS, Dimitrios; VICEN-BUENO, Raul. Sensors to increase the security of underwater communication cables: A review of underwater monitoring sensors. *Sensors*, v. 20, n. 3, 2020. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/20/3/737>>. 2.11.1
- ESQUADRA, Ilques Barbosa Junior. Almirante de. *No mar, a riqueza e o futuro das nações*. 2019. <https://www.marinha.mil.br/economia-azul/noticias/no-mar-riqueza-e-o-futuro-das-nacoes>. Accessed 2023-01-12. 1.3
- FABBRI, Cameron; ISLAM, Md Jahidul; SATTAR, Junaed. Enhancing underwater imagery using generative adversarial networks. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. [S.l.: s.n.], 2018. p. 7159–7165. 2.4.1, 2.4.2
- FERNANDO, Dr. *A visão dos cães: uma maneira diferente de ver o mundo*. 2018. Url=<https://www.sahv.com.br/post/a-visao-dos-caes-uma-maneira-diferente-de-ver-o-mundo/130>. Accessed: 2023-02-15. 2.12
- FLORES, Denisele. *Sinapses*. 2020. Url = <https://escolaeducacao.com.br/sinapses/>. Accessed: 2023-02-16. 2.17
- G1. *Navio russo suspeito de espionagem coloca Marinha Brasileira em alerta*. 2020. <<https://epocanegocios.globo.com/Brasil/noticia/2020/02/>>

[epoca-negocios-navio-russo-suspeito-de-espionagem-coloca-marinha-brasileira-em-alerta.html](#)>. Accessed: 2023-03-15. 2.11.1

_____. *5 cidades ao redor do mundo que estão debaixo d'água*. 2022. <https://g1.globo.com/mundo/noticia/2022/01/30/5-cidades-ao-redor-do-mundo-que-estao-debaixo-dagua.ghtml>. Accessed 2023-02-26. 2.11.3

GHAFOOR, Huma; NOH, Youngtae. An overview of next-generation underwater target detection and tracking: An integrated underwater architecture. *Ieee Access*, IEEE, v. 7, p. 98841–98853, 2019. 2.4.1, 2.4.1, 2.4.1

GIRSHICK, Ross. Fast r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. 2.9

GIRSHICK, Ross *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 580–587. (document), 2.9, 2.38

GONZALEZ, R. C.; WOODS R. E. *Processamento digital de imagens. 3. ed.* [S.l.]: Pearson Prentice Hall, 2010. 2.6, 2.3

HAN, J.; KAMBER, M. *Data Mining: Concepts and Techniques*. [S.l.]: Morgan Kaufmann, 2011. ISBN 978-9380931913. 4

HASSAN, Wan Hafiza Wan *et al.* Experimental study of light wave propagation for underwater optical wireless communication (uowc). *J. Commun.*, v. 17, n. 1, p. 23–29, 2022. 2.4.2

HAYKIN, S. *Redes Neurais: Princípios e Prática*. [S.l.]: Bookman Editora, 2001. ISBN 9788577800865. (document), 2.19, 2.6, 2.6, 2.6, 2.7, 2.31

HEATON, Jeff. *Introduction to Neural Networks for Java*. [S.l.]: Heaton Research, 2008. ISBN 978-1604390087. 4

HEBB, Donald Olding. *The organization of behavior: A neuropsychological theory*. [S.l.]: Psychology Press, 2005. 2.6

HIRA, Samed. *ACTIVATION FUNCTIONS*. 2021. <<https://samedhira.medium.com/activation-functions-cf6fef0e5922>>. Accessed: 2023-03-09. 2.23

HOOKER, Peter. “to die gallantly”?—the role of the surface fleet in german naval strategy, 1919–41,“. *Naval War College*, v. 74, n. 4, 2021. 2.4.2

HOŻYŃ, Stanisław. A review of underwater mine detection and classification in sonar imagery. *Electronics*, MDPI, v. 10, n. 23, p. 2943, 2021. 2.11.5

HUO, Guanying; WU, Ziyin; LI, Jiabiao. Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data. *IEEE Access*, v. 8, p. 47407–47418, 2020. 2.10

_____. _____. *IEEE access*, IEEE, v. 8, p. 47407–47418, 2020. 1

JANA, Abhisek. *Understand and Implement the Backpropagation Algorithm From Scratch In Python*. 2019. <<http://www.adeveloperdiary.com/data-science/machine-learning/understand-and-implement-the-backpropagation-algorithm-from-scratch-in-python/>>. Accessed: 2023-02-25. (document), 2.29

- JIANG, Xudong. Feature extraction for image recognition and computer vision. In: *2009 2nd IEEE International Conference on Computer Science and Information Technology*. [S.l.: s.n.], 2009. p. 1–15. [2.3](#)
- JIN, Leilei; LIANG, Hong. Deep learning for underwater image recognition in small sample size situations. In: *OCEANS 2017 - Aberdeen*. [S.l.: s.n.], 2017. p. 1–4. [2.10](#)
- JIN, Leilei; LIANG, Hong; YANG, Changsheng. Accurate underwater atr in forward-looking sonar imagery using deep convolutional neural networks. *IEEE Access*, IEEE, v. 7, p. 125522–125531, 2019. [2.10](#)
- KALLASI, F. *et al.* Computer vision in underwater environments: a multiscale graph segmentation approach. In: *OCEANS*. [S.l.: s.n.], 2015. p. 1–6. [2](#)
- KEZEBOU, Landry *et al.* Underwater object tracking benchmark and dataset. In: IEEE. *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*. [S.l.], 2019. p. 1–6. [3](#)
- KHALEDI, Siamak *et al.* Design of an underwater mine detection system. *2014 Systems and Information Engineering Design Symposium (SIEDS)*, p. 78–83, 2014. Disponível em: <https://api.semanticscholar.org/CorpusID:10707436>. [2.11.5](#)
- KONEN, Christina S; KASTNER, Sabine. Two hierarchically organized neural systems for object information in human visual cortex. *Nature neuroscience*, Nature Publishing Group, v. 11, n. 2, p. 224–231, 2008. [2.2](#)
- LAVRENKO, Victor. *Neural Networks 6: solving XOR with a hidden layer*. 2016. https://www.youtube.com/watch?v=kNPGXgzxoHw&t=7s&ab_channel=VictorLavrenko. Accessed 2023-03-28. (document), [2.27](#)
- LEE, Sejin; PARK, Byungjae; KIM, Ayoung. Deep learning based object detection via style-transferred underwater sonar images. *IFAC-PapersOnLine*, Elsevier, v. 52, n. 21, p. 152–155, 2019. [2.4.1](#), [2.10](#)
- LIEBE, Stefanie *et al.* Theta coupling between v4 and prefrontal cortex predicts visual short-term memory performance. *Nature neuroscience*, Nature Publishing Group, v. 15, n. 3, p. 456–462, 2012. [2.2](#)
- MALVA, Pamela. *A CIDADE SUBMERSA DE HERACLEION, UMA DAS DESCOBERTAS MAIS IMPORTANTES DO EGITO ANTIGO*. 2020. <https://aventurasnahistoria.uol.com.br/noticias/galeria/historia-a-cidade-submersa-de-heracleion-egito-antigo.phtml>. Accessed 2023-02-26. (document), [2.11.3](#), [2.42](#)
- MATOS, Aníbal *et al.* Multiple robot operations for maritime search and rescue in eurathlon 2015 competition. In: *OCEANS 2016 - Shanghai*. [S.l.: s.n.], 2016. p. 1–7. [2.11.4](#)
- MATOS, David. *Customizando Redes Neurais com Funções de Ativação Alternativas*. 2022. <https://www.cienciaedados.com/customizando-redes-neurais-com-funcoes-de-ativacao-alternativas/>. Accessed: 2023-03-08. [2.20](#), [2.21](#)
- MCCULLOCH, Warren S; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, p. 115–133, 1943. [2.6](#)

- MORONI, D *et al.* Underwater manmade and archaeological object detection in optical and acoustic data. *Pattern recognition and image analysis*, Springer, v. 24, n. 2, p. 310–317, 2014. [1.1](#), [2.11.3](#)
- MURPHY, AJ; LANDAMORE, MJ; BIRMINGHAM, RW. The role of autonomous underwater vehicles for marine search and rescue operations. *Underwater Technology*, Society for Underwater Technology, v. 27, n. 4, p. 195–205, 2008. [2.11.4](#)
- NADIMI, Nahid; JAVIDAN, Reza; LAYEGHI, Kamran. Efficient detection of underwater natural gas pipeline leak based on synthetic aperture sonar (sas) systems. *Journal of Marine Science and Engineering*, MDPI, v. 9, n. 11, p. 1273, 2021. [1.1](#)
- NAVAL, Escola de Guerra. *EMA-305 - Doutrina Básica da Marinha*. Rio de Janeiro, 2003. Confidencial. [2.11.5](#)
- NEUPANE, Dhiraj; SEOK, Jongwon. A review on deep learning-based approaches for automatic sonar target recognition. *Electronics*, MDPI, v. 9, n. 11, p. 1972, 2020. [2.4.2](#), [2.5](#)
- OLEARI, F. *et al.* An underwater stereo vision system: from design to deployment and dataset acquisition. In: *OCEANS*. [S.l.: s.n.], 2015. p. 1–5. [2](#)
- O'SHEA, Keiron; NASH, Ryan. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015. [2.8](#)
- PADILHA, Luiz. *Tesouro azul no Atlântico*. 2020. <<https://www.defesaaereanaval.com.br/naval/tesouro-azul-no-atlantico>>. Accessed: 2023-03-03. ([document](#)), [2.41](#)
- PAL, Kaushika; PATEL, Biraj. V. Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques. In: *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*. [S.l.: s.n.], 2020. p. 83–87. [3.3](#)
- PALERMO, Celso. *Gerenciamento de cores para fotógrafos: Como funciona a captura digital*. [S.l.]: IAR - UNICAMP, 2015. [2.5](#), [2.3](#)
- PANETTA, Karen *et al.* Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with gan. *IEEE Journal of Oceanic Engineering*, v. 47, n. 1, p. 59–75, 2022. [3](#)
- PITERI, Marco Antônio; RODRIGUES, José Carlos. Fundamentos de visão computacional. *Presidente Prudente: FCT/UNESP-PP*, 2011. [2.3](#)
- PRETOLÂNDIA-PE, Prefeitura de. *PETROLÂNDIA VELHA / ANTIGA PETROLÂNDIA*. 2017. [Http://www.petrolandia.pe.gov.br/historia](http://www.petrolandia.pe.gov.br/historia). Accessed 2023-07-10. [2.11.3](#)
- RAMCHOUN, Hassan *et al.* Multilayer perceptron: Architecture optimization and training. *International Journal of Interactive Multimedia and Artificial Intelligence ...*, 2016. ([document](#)), [2.28](#), [2.7](#)
- REDMON, Joseph *et al.* You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788. ([document](#)), [2.9](#), [2.39](#)

- RIZZINI, D. Lodi *et al.* Investigation of vision-based underwater object detection with multiple datasets. *IJARS, InTech*, -, n. -, p. -, may 2015. [2](#)
- ROCHA, Bárbara. *Neurônios – glossário do cérebro*. 2016. <https://metodosupera.com.br/neuronios-glossario-do-cerebro/>. Accessed: 2022-08-13. ([document](#)), [2.16](#)
- ROSENBLATT, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. [2.18](#), [2.6](#)
- ROSLI, Mohamed Syazwan Asyraf Bin *et al.* Underwater animal detection using yolov4. In: IEEE. *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSC)*. [S.l.], 2021. p. 158–163. [1.1](#)
- SCHINDELIN, Johannes *et al.* The imagej ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, v. 82, n. 7-8, p. 518–529, 2015. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mrd.22489>. [3](#)
- SHAPIRO, Linda G; STOCKMAN, G. *Computer vision*. [S.l.]: Michigan State University, 2000. [2.3](#)
- SIDDHARTHA, Jetty Bangaru; JAYA, T.; RAJENDRAN, V. Rdnm for classification and prediction of rock/mine in underwater acoustics. *Materials Today: Proceedings*, 2021. ISSN 2214-7853. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2214785321050756>. [1.1](#)
- SILVEIRA, Igor Vilas-Bôas *et al.* A multi-layer perceptron model for underwater object recognition. *JOURNAL OF BIOENGINEERING OF AND TECCHNOLOGY applied to Health*, v. 5, n. 4, p. 266–271, 2022. [2.1](#)
- SUBSEA, DOF. *Glider AUV*. 2018. [Url=https://www.dofsubsea.com/rov/glider-auv/](https://www.dofsubsea.com/rov/glider-auv/). Accessed: 2023-02-15. [2.11](#)
- SUN, Kai; CUI, Weicheng; CHEN, Chi. Review of underwater sensing technologies and applications. *Sensors*, MDPI, v. 21, n. 23, p. 7849, 2021. [2.4.2](#)
- TENG, Bowen; ZHAO, Hongjian. Underwater target recognition methods based on the framework of deep learning: A survey. *International Journal of Advanced Robotic Systems*, SAGE Publications Sage UK: London, England, v. 17, n. 6, p. 1729881420976307, 2020. [2.4.2](#), [2.10](#), [2.11.5](#)
- TRUVER, Scott C. Taking mines seriously: Mine warfare in china’s near seas. *Naval War College Review*, JSTOR, v. 65, n. 2, p. 30–66, 2012. ([document](#)), [2.45](#)
- VENN, Fjodor Van. *The Neural Network Zoo*. 2016. <https://www.asimovinstitute.org/neural-network-zoo/>. Accessed: 2023-03-09. ([document](#)), [2.24](#)
- YU, Xian *et al.* Man-made object recognition from underwater optical images using deep learning and transfer learning. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2018. p. 1852–1856. [2.10](#)

Artigo completo publicado em periódico

A Multi-Layer Perceptron Model for Underwater Object Recognition

Igor Vilas-Bôas Silveira

Roberto Monteiro

Oberdan Pinheiro Rocha

Alex Álisson Bandeira Santos

DOI: <https://doi.org/10.34178/jbth.v5i4.250>

Keywords: Computer Model, Underwater Objects, Classification, Recognition, Accuracy

Abstract

The more human gets interested in sea exploration, the more research to detect objects under the water is done. Therefore, the ability to detect, classify, recognize and track all kinds of objects is evolving daily. This paper aims to introduce a computer model for underwater object classification and recognition based on a Multilayer Perceptron network. The model was constructed with a mixed dataset for the training phase, combining artificial and natural objects, and it reached approximately 99.97% classifying accuracy.

Modelo computacional para reconhecimento de objetos em ambiente subaquático baseado em aprendizagem de máquina

Igor Vilas-Bôas Silveira

Salvador, Novembro de 2023.