



William Alberto Ramirez Ruiz

Aplicação e avaliação do desempenho de métodos de aprendizado profundo para classificação de imagens de radar SAR (synthetic aperture radar) para monitoramento de áreas marinhas na detecção de feições de interesse para a área de óleo e gás

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia elétrica, do Departamento de Engenharia elétrica da PUC-Rio.

Orientador : Prof. Marco A. Cavalcanti Pacheco
Coorientador: Dr. Pedro M. Achanccaray Diaz

Rio de Janeiro
Abril de 2021



William Alberto Ramirez Ruiz

Aplicação e avaliação do desempenho de métodos de aprendizado profundo para classificação de imagens de radar SAR (synthetic aperture radar) para monitoramento de áreas marinhas na detecção de feições de interesse para a área de óleo e gás

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia elétrica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

Prof. Marco A. Cavalcanti Pacheco

Orientador

Departamento de Engenharia elétrica – PUC-Rio

Dr. Pedro M. Achanccaray Diaz

Coorientador

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Dra. Manoela Rabello Kohler

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Prof. Leonardo Alfredo Forero Mendoza

UERJ

Dra. Cristina Maria Bentz

PETROBRAS

Dr. Jefersson Alex Dos Santos

UFMG

Rio de Janeiro, 30 de Abril de 2021

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

William Alberto Ramirez Ruiz

O autor formou-se em Engenharia Eletrônica pela Universidade Antonio Nariño (UAN) em 2017. Seu trabalho é focado em processamento digital de imagens, sensoriamento remoto e Deeplearning.

Ficha Catalográfica

Ramirez Ruiz, William Alberto

Aplicação e avaliação do desempenho de métodos de aprendizado profundo para classificação de imagens de radar SAR (synthetic aperture radar) para monitoramento de áreas marinhas na detecção de feições de interesse para a área de óleo e gás / William Alberto Ramirez Ruiz; orientador: Marco A. Cavalcanti Pacheco; coorientador: Pedro M. Achanccaray Diaz. – 2021.

94 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia elétrica, 2021.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. CNN. 3. Sensoriamento remoto. 4. Imageamento por Radar. 5. Derramamento de óleo. 6. SAR. I. Marco A. Cavalcanti Pacheco. II. Pedro M. Achanccaray Diaz. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia elétrica. IV. Título.

CDD: 621.3

Aos meus pais que me apoiaram e
me incentivaram para sempre seguir meu caminho.

Agradecimentos

Quero agradecer ao meu orientador e co-orientador pelo apoio durante a execução da minha dissertação.

Agradeço aos meus pais, William e Jaqueline, pelo apoio constante nesta fase da minha vida. Suas palavras e conselhos me ajudaram a superar momentos difíceis, permitindo-me alcançar meus objetivos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Ramirez Ruiz, William Alberto; Marco A. Cavalcanti Pacheco; Pedro M. Achanccaray Diaz. **Aplicação e avaliação do desempenho de métodos de aprendizado profundo para classificação de imagens de radar SAR (synthetic aperture radar) para monitoramento de áreas marinhas na detecção de feições de interesse para a área de óleo e gás.** Rio de Janeiro, 2021. 94p. Dissertação de Mestrado – Departamento de Engenharia elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O estudo dos eventos naturais e dos gerados pela atividade humana no mar tem tido uma grande prioridade para o setor de petróleo, isso devido à possibilidade de ter um evento perigoso para o ambiente marinho ou a área de produção. Nesse contexto, o objetivo deste trabalho é a avaliação de abordagens baseadas em aprendizado profundo para a classificação de eventos no mar usando imagens de radar de abertura sintética na área de óleo e gás. Métodos baseados em aprendizado profundo têm mostrado um ótimo desempenho através do uso de camadas convolucionais, onde as características são extraídas automaticamente através da definição de um *kernel* e *stride*. As seguintes arquiteturas são avaliadas neste trabalho: Inception V3, Xception, Inception ResNet V2, MobileNet, VGG16 e Deep Attention sampling. A avaliação é feita em uma metodologia de classificação de eventos no mar usando duas bases de dados de imagens de radar: a primeira contém 10 eventos comumente presentes no oceano ártico, e a segunda descreve um derramamento de óleo presente na costa da Louisiana. Nos experimentos realizados se obteve os melhores resultados com as arquiteturas Deep Attention sampling as quais atingiram valores de f1-score e *Recall* de até 0.82% e 0.87% respectivamente, para a classe de interesse no conjunto de dados de derramamento de óleo. Para o conjunto de dados de eventos naturais no mar, um alto desempenho foi evidenciado para arquiteturas baseadas no uso de módulos de Inception, tendo pontuações mais altas de F1-score e *Recall* para a arquitetura Xception. Além disso, foi observado uma melhoria de até 10% e 13% nas métricas f1-score e *Recall* no uso da atenção, em relação à sua arquitetura base (VGG16), e 4% respeito a outras arquiteturas baseadas em módulos Inception, isto para o conjunto de dados de eventos no mar, demonstrando as vantagens de usar amostragem com atenção.

Palavras-chave

CNN; Sensoriamento remoto; Imageamento por Radar; Derramamento de óleo; SAR.

Abstract

Ramirez Ruiz, William Alberto; Marco A. Cavalcanti Pacheco (Advisor); Pedro M. Achancaray Diaz (Co-Advisor). **Application and performance evaluation of deep learning methods for classification of SAR (synthetic aperture radar) radar images for monitoring marine areas in detecting features of interest to the oil and gas area.** Rio de Janeiro, 2021. 94p. Dissertação de Mestrado – Departamento de Engenharia elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The study of natural events and those generated by human activity at sea has been a high priority for the Oil & Gas industry, due to the possibility of a dangerous event for the marine environment or the production area. In this context, the objective of this work is the evaluation of approaches based on deep learning for the classification of events in the sea using synthetic aperture radar images in the oil and gas area. Methods based on deep learning have shown an excellent performance through the use of convolutional layers, where the characteristics are extracted automatically through the definition of a kernel and stride. The following architectures are evaluated in this work: Inception V3, Xception, Inception ResNet V2, MobileNet, VGG16, and Deep Attention sampling. The assessment is made using a methodology for classifying events at sea using two radar image databases: the first contains 10 events commonly present in the Arctic Ocean, and the second describes an oil spill present near the Louisiana coast. In the experiments performed, the best results were obtained with the Deep Attention sampling architectures, which reached f1-score and Recall values of up to 0.82% and 0.87% respectively, for the class of interest in the oil spill dataset. For the dataset of natural events in the sea, high performance was evidenced for architectures based on the non-use of Inception modules, having higher values of F1-score and Recall for an Xception architecture. Also, an improvement of up to 10% and 13% in the metrics f1-score and recall in the use of attention was observed, concerning its base architecture (VGG16), and 4% with other architectures based on Inception modules, this for the dataset of events at sea, demonstrating the advantages of using Attention Sampling carefully.

Keywords

CNN; Remote sensing; Imaging Radar; Oil spill; SAR.

Sumário

1	INTRODUÇÃO	16
1.1	Motivação	16
1.2	Objetivos	18
1.2.1	Objetivo geral	18
1.2.2	Objetivos específicos	18
1.3	Contribuições	18
1.4	Organização dos capítulos restantes da dissertação	19
2	TRABALHOS RELACIONADOS	20
3	FUNDAMENTOS	25
3.1	Sensoriamento Remoto	25
3.2	Radar	27
3.2.1	Radar de Abertura Real	27
3.2.2	Radar de Abertura Sintética	29
3.2.3	Características dos sistemas SAR	31
3.3	Radar para monitoramento de áreas marinhas	36
4	REDES NEURAIS CONVOLUCIONAIS PARA CLASSIFICAÇÃO	39
4.1	Conceitos básicos	39
4.1.1	Camadas convolucionais	40
4.1.2	Camada de Pooling	41
4.1.3	Camada de Dropout	41
4.1.4	Camada totalmente conectada	42
4.2	Métricas de avaliação	42
4.2.1	Matriz de confusão	42
4.2.2	Precision	43
4.2.3	Recall	43
4.2.4	F1-score	43
4.2.5	GradCAM	43
4.3	Arquiteturas estudadas	44
4.3.1	VGG16	44
4.3.2	Inception V3 (Incep)	45
4.3.3	Xception (Xcep)	46
4.3.4	Inception ResNet v2 (Incep-R)	48
4.3.5	MobileNet (MNet)	50
4.3.6	Deep Attention Sampling (DAS)	52
5	METODOLOGIA	54
5.1	Pré-processamento dados Sentinel-1A	54
5.2	Construção do conjunto de dados	55
5.3	Implementação de modelos baseados em CNN	56
5.3.1	Sintonização de hiperparâmetros	57

5.3.2	Processo de Inferência	58
6	ANÁLISE EXPERIMENTAL	59
6.1	Conjuntos de Dados	59
6.1.1	Eventos Naturais no Mar	59
6.1.2	Derramamento de Óleo no Mar	62
6.2	Configuração Experimental	64
6.3	Protocolo experimental	66
6.4	Resultados	67
6.4.1	Eventos Naturais no Mar	67
6.4.2	Derramamento de Óleo no Mar	72
7	CONCLUSÕES	80
	Referências bibliográficas	82
	Apêndices	86
A	Resultados de Inferência do conjunto de dados de Derramamento de Óleo no mar.	86
B	Resultados de Inferência do conjunto de dados de eventos naturais no mar.	90

Lista de figuras

Figura 1.1	Processo de expansão do óleo no derramamento no Golfo do México no ano 2010, obtido de [3].	17
Figura 3.1	Resolução radial no processo de medição da radiação do espectro eletromagnético utilizando Radar.	27
Figura 3.2	Resolução azimute no processo de medição da radiação do espectro eletromagnético utilizando Radar.	28
Figura 3.3	Processo de sensoramento usando radar de abertura sintética, obtido de [21].	29
Figura 3.4	Polarização VH vs polarização VV para detecção de derramamento de óleo, produto de Sentinel-1A Level 1.	31
Figura 3.5	Efeito das bandas de frequência no mapeamento usando Radar.	32
Figura 3.6	Processo de detecção baseado no ângulo de incidência, mostrando a geração de áreas com encurtamento.	33
Figura 3.7	Distorções na imagem SAR: inversão e encurtamento. Obtido de [20]	34
Figura 3.8	Espalhamento do pulso emitido com base nos tipos de superfícies, obtido de [20].	34
Figura 3.9	Alteração na imagem do radar a partir da variação da rugosidade da área de estudo, obtido de [20].	35
Figura 3.10	Banda C vs Banda L mapeamento no mar (Gelo marinho), obtido de [22].	37
Figura 4.1	LeNet-5. obtido de [23].	40
Figura 4.2	Exemplo da operação de Convolução.	40
Figura 4.3	Operação de Max Pooling: imagem 4x4, <i>kernel</i> 2x2, <i>stride</i> = 2.	41
Figura 4.4	Operação de Dropout: 50% de Dropout.	42
Figura 4.5	Matriz de confusão: cálculo dos verdadeiros positivos(TP), verdadeiros negativos(TN), falsos positivos(FP) e falsos negativos(FN) para uma classe prevista.	42
Figura 4.6	Modulo de VGG16, obtido de [26].	44
Figura 4.7	Fatoração de convoluções 5×5 em convoluções menores 3×3 , obtido de [27].	45
Figura 4.8	Inception original vs módulo de Inception proposto (<i>Grid</i> 8×8), obtido de [27].	46
Figura 4.9	Implementação do classificador auxiliar como regularizador: O Classificador auxiliar é usado na parte superior da última camada 17×17 , tendo um Batch Normalization, ganho Top-1 error 0,4% , obtido de [27].	46
Figura 4.10	Modulo de Xception, obtido de [28].	47
Figura 4.11	Bloco Inception com conexões residuais, obtido de [29].	48
Figura 4.12	Dimensionamento dos resíduos usando uma ativação com escalonamento, obtido de [29].	49

Figura 4.13	Convolução separável em profundidade, obtido de [30] .	50
Figura 4.14	Amostragem por atenção usando uma rede totalmente convolucional, obtido de [19].	52
Figura 5.1	Metodologia para gerar os mapas de classificação usando modelos baseados em aprendizado profundo.	54
Figura 5.2	Grafo de pré-processamento para as imagens Sentinel-1 do conjunto de dados de derramamento de óleo.	55
Figura 5.3	Contexto com base na quantidade de informações em um tamanho de <i>patch</i> .	56
Figura 5.4	Processo de inferência usando <i>patches</i> .	58
Figura 6.1	Localização da área de estudo para o conjunto de dados de Eventos Naturais no Mar.	60
Figura 6.2	Imagens das classes presentes no conjunto de dados de Eventos Naturais no Mar, (a) Ondas puras do oceano, (b) Rajada de vento, (c) Células micro-convectivas, (d) Células de chuva, (e) Manchas biológicas, (f) Gelo marinho, (g) Icebergs, (h) Área de vento fraco, (i) Frente atmosférica, and (j) Frente oceânica, obtido de [6]	61
Figura 6.3	Área de interesse para o conjunto de dados de Derramamento de Óleo no Mar.	63
Figura 6.4	Conjunto de relatórios obtidos na NOAA para eventos no mar do Golfo do México. Os relatórios mostram o vetor da anotação junto com as coordenadas, órbita e data da imagem de radar estudada.	63
Figura 6.5	<i>Patches</i> gerados pelo módulo de atenção. Os <i>patches</i> foram geradas a partir da estimativa de Monte Carlo considerando uma atenção espacial, com base na derivação do gradiente.	66
Figura 6.6	Curvas de treinamento no conjunto de dados de Eventos Naturais no Mar: F1-score vs. Número de Épocas. O gráfico apresenta a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.	67
Figura 6.7	Intervalo de confiança das curvas de treinamento no conjunto de dados de Eventos Naturais no Mar: F1-score vs. Épocas. O gráfico apresenta o intervalo de confiança para a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.	68
Figura 6.8	GradCAM dos modelos avaliados no conjunto de dados de Eventos Naturais no Mar.	71
Figura 6.9	Curvas de treinamento no conjunto de dados de derramamento de óleo: F1-score vs. Épocas. O gráfico apresenta a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.	73
Figura 6.10	Intervalo de confiança das curvas de treinamento no conjunto de dados do Derramamento de Óleo no Mar: F1-score vs. Número de Épocas. O gráfico apresenta o intervalo de confiança para a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.	74

Figura 6.11 Resultado da inferência de uma imagem SAR (19/12/2018) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de <i>patches</i> , respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).	76
Figura 6.12 Resultado da inferência de uma imagem SAR (15/10/2019) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de <i>patches</i> , respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).	77
Figura 6.13 <i>patches</i> das imagens SAR de teste no conjunto de dados de Derramamento de Óleo no Mar. O GradCAM apresenta o mapa de ativação para inferência, onde a matriz de ativação é sobreposta na imagem do estudo.	78
Figura A.1 Resultado da inferência de uma imagem SAR (24/02/2020) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).	86
Figura A.2 Resultado da inferência de uma imagem SAR (19/03/2020) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).	87
Figura A.3 Resultado da inferência de uma imagem SAR (19/03/2018) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).	88
Figura A.4 Resultado da inferência de uma imagem SAR (19/12/2018) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).	89
Figura B.1 GradCAM para amostras da classe Ondas puras do oceano.	90
Figura B.2 GradCAM para amostras da classe Rajadas de vento.	90
Figura B.3 GradCAM para amostras da classe Células Microconvectivas.	91
Figura B.4 GradCAM para amostras da classe Manchas biológicas.	91
Figura B.5 GradCAM para amostras da classe Células de chuva.	92
Figura B.6 GradCAM para amostras da classe Gelo marinho.	92
Figura B.7 GradCAM para amostras da classe Iceberg.	93
Figura B.8 GradCAM para amostras da classe Áreas de vento fraco.	93
Figura B.9 GradCAM para amostras da classe Frente atmosférica.	94
Figura B.10 GradCAM para amostras da classe Frente oceânica.	94

Lista de tabelas

Tabela 3.1	Plataformas usadas no sensoriamento remoto.	26
Tabela 3.2	Parâmetros de aquisição no uso de radar e descrição do terreno de estudo.	29
Tabela 3.3	Bandas de frequência para SAR.	32
Tabela 4.1	Top-1 e Top-5 da métrica Precision no conjunto de dados ImageNet, obtido de [28].	48
Tabela 4.2	Top-1 e Top-5 do erro para as arquiteturas baseadas em Inception no desafio ILSVRC 2012, obtido de [29].	50
Tabela 6.1	Distribuição do conjunto de imagens para Eventos Naturais no Mar, dados obtidos próximo ao oceano Antártico, Índico e Arábico no 2016.	60
Tabela 6.2	Conjunto de dados de Radar para derramamento de óleo, obtidos entre 2018-2020. As imagens em negrito são usadas para treinamento e as outras para teste.	64
Tabela 6.3	<i>Patches</i> gerados do conjunto de dados de Derramamento de óleo no Mar considerando um tamanho de 512×512 e <i>stride</i> 10%.	64
Tabela 6.4	<i>Patches</i> gerados do conjunto de dados de Derramamento de óleo no Mar considerando um tamanho de 256×256 e <i>stride</i> 10%.	65
Tabela 6.5	Arquiteturas avaliadas na classificação de eventos no mar.	65
Tabela 6.6	Parâmetros de treinamento.	66
Tabela 6.7	Resultados da métrica <i>Recall</i> dos modelos avaliados no conjunto de teste das imagens de Eventos naturais no Mar.	69
Tabela 6.8	Resultados da métrica F1-score dos modelos avaliados no conjunto de teste das imagens de Eventos naturais no Mar.	70
Tabela 6.9	Resultados da métrica F1 score dos modelos avaliados no conjunto de teste das imagens de Derramamento de Óleo no Mar.	75
Tabela 6.10	Resultados da métrica Recall dos modelos avaliados no conjunto de teste das imagens de Derramamento de Óleo no Mar.	75
Tabela 6.11	Resultados da métrica Recall no conjunto de teste com dados confidenciais	75

Lista de Abreviaturas

ANN – Rede neural artificial (Artificial Neural Network)

CNN – Redes neurais convolucionais (Convolutional Neural Networks)

DAS – Deep Attention sampling

EMS – Agência Europeia de Segurança Marítima

ESA – Agência Espacial Europeia

EW – Extra Wide swath

GRD – Ground Range Detected

GLCMs – matrizes de co-ocorrência em níveis de cinza

IW – Interferometric Wide

Incep – Inception V3

Incep-R – Inception ResNet V2

MNet – MobileNet

NOAA – Administração Oceânica e Atmosférica Nacional

SVM – Vetores de suporte (Support Vector Machines)

RAR – Radar de Abertura Real (Real Aperture Radar)

SAR – Radar de Abertura Sintética (Synthetic Aperture Radar)

SM – Stripmap

SLC – Single Look Complex

WV – WaVe mode

Xcep – Xception

Nossos primeiros esforços são puramente instintiva, sussurros de uma imaginação vívida e indisciplinado

Nikola Tesla, *Minhas Invenções A Autobiografia de Nikola Tesla (1856-1943) Nikola Tesla & Ben Johnston.*

1 INTRODUÇÃO

1.1 Motivação

No mar ocorrem diferentes eventos os quais podem ser naturais ou gerados pelo homem. Entre os eventos naturais como exemplo, temos o impacto de altas temperaturas na superfície da água ou a presença de ventos fortes. Por outro lado, os eventos feitos pelo homem surgem pelo interesse em aproveitar os recursos naturais do mar. Estes eventos tem um alto impacto no meio ambiente, como é o caso do derramamento de óleo [1, 2], que pode levar a perdas de espécies animais junto com a contaminação de grandes extensões da flora marinha. Devido a isso tornou-se necessário monitorar a superfície do mar na busca de eventos que podem representar um risco potencial para a preservação do meio ambiente.

Entre os eventos feitos pelo homem mais críticos para o meio ambiente temos o derramamento de óleo devido à sua rápida propagação ao interagir com outros eventos climáticos comuns no mar. Um exemplo claro é o derramamento de óleo no Golfo de México no ano 2010, o qual mostrou o forte impacto que esse tipo de evento tem na natureza, e como os fluidos podem ser incontroláveis se não são tomadas medidas rápidas [3]. Para a contenção destes tipos de eventos é necessário mapeamento rápido da área afetada, para seu posterior isolamento e neutralização do óleo na área do incidente. Na Figura 1.1 podemos ver a rápida expansão do óleo para outras áreas próximas, tornando-se quase incontrolável se as medidas corretas não forem tomadas, gerando danos no longo prazo.

Existem sistemas de monitoramento do mar que empregam diferentes plataformas para a detecção deste tipo de eventos: Drones [4], Aviões [5], e Satélites [6]. O monitoramento usando satélites é o mais usado para uma detecção rápida dos eventos devido a sua grande cobertura, custo razoável e altas resoluções espaciais e temporais. Sensores ativos (e.g. radar) são preferidos em vez de sensores passivos (e.g. óticos), pois eles quase não são afetados pelas condições meteorológicas (presença de nuvens ou tempestades) e podem adquirir imagens durante o dia e a noite.

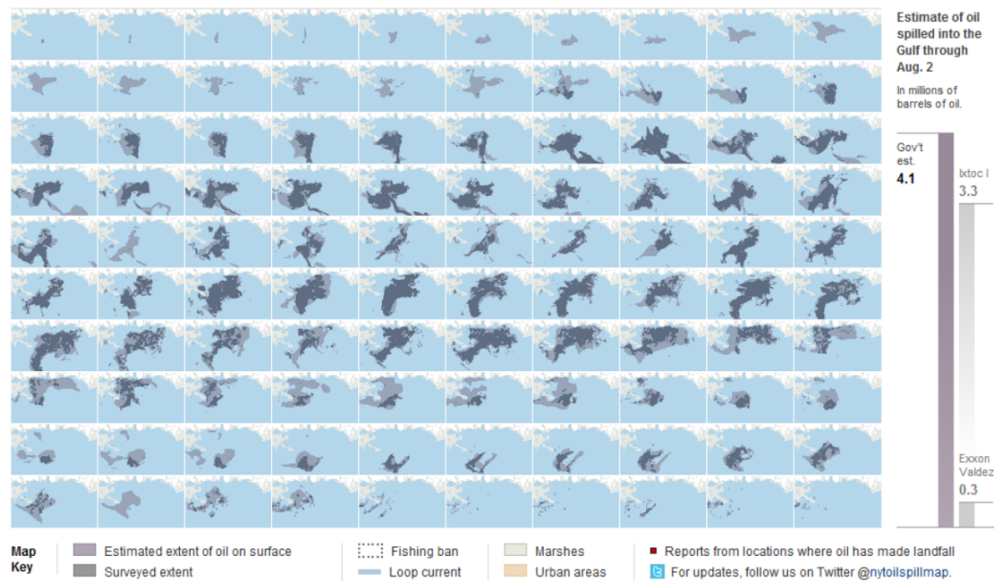


Figura 1.1: Processo de expansão do óleo no derramamento no Golfo do México no ano 2010, obtido de [3].

Os satélites de Radar de Abertura Sintética (Synthetic Aperture Radar - SAR) são um tipo de sensores ativos usados com sucesso para o monitoramento do mar. Por exemplo, em [7] é mostrado como o uso do Radar pode fornecer uma grande quantidade de informações em áreas marítimas, descrevendo em detalhe a rugosidade, forma e características relacionadas à retroespalhamento do óleo.

Nesse contexto o presente trabalho tem como objetivo avaliar o desempenho de métodos de aprendizado profundo para a classificação de eventos no mar usando imagens de radar de abertura sintética (SAR). Os métodos de aprendizado profundo [6, 8] destacam-se por seu ótimo desempenho em cenários de alta complexidade superando os métodos tradicionais baseados em algoritmos de classificação como Máquinas de Vetores de Suporte (SVM) [10, 11] ou Random Forest [7], que requerem um processo de extração de atributos desenhado por um especialista. Por outro lado, no aprendizado profundo é feito um aprendizado *end-to-end*, onde os atributos mais representativos são aprendidos e extraídos pelo modelo, apresentando um grande benefício na automatização do estudo de atributos, havendo uma menor intervenção de especialistas.

1.2

Objetivos

1.2.1

Objetivo geral

O objetivo principal deste trabalho é aplicar e avaliar o desempenho de métodos de aprendizado profundo para classificação de imagens de monitoramento de áreas marinhas na detecção de feições de interesse para a área de óleo e gás.

1.2.2

Objetivos específicos

Os objetivos específicos são apresentados a seguir:

1. Estudar métodos estado da arte baseados em aprendizado profundo para classificação de imagens.
2. Aplicar métodos estado da arte baseados em aprendizado profundo para classificação de derramamentos de óleo e diferentes eventos no mar.
3. Avaliar de forma quantitativa e qualitativa métodos estado da arte baseados em aprendizado profundo.
4. Comparar métodos estado da arte baseados em aprendizado profundo para classificação de derramamentos de óleo e diferentes eventos no mar.

1.3

Contribuições

As principais contribuições deste trabalho são apresentadas a seguir:

1. Uma revisão bibliográfica sobre métodos para classificação de imagens aplicados ao monitoramento do mar.
2. Um repositório público com as implementações dos métodos estado da arte baseados em aprendizado profundo para classificação de imagens de monitoramento do mar¹.
3. Uma avaliação quantitativa e qualitativa dos métodos estado da arte baseados em aprendizado profundo para classificação de imagens de monitoramento do mar.

¹https://github.com/williamalbert94/Oil_spill_classification

1.4

Organização dos capítulos restantes da dissertação

O Capítulo 2 descreve os trabalhos relacionados à detecção de eventos no mar usando imagens de radar com foco nos abordagens baseadas em redes neurais convolucionais.

O Capítulo 3 apresenta os fundamentos teóricos relacionados a sensoriamento remoto, princípio de funcionamento do radar e redes neurais convolucionais.

O Capítulo 4 explica os abordagens avaliados nesta dissertação.

O Capítulo 5 descreve a metodologia empregada para a classificação de eventos no mar.

O Capítulo 6 mostra os conjuntos de dados usados, o protocolo experimental e os resultados obtidos nos experimentos.

Finalmente, o Capítulo 7 apresenta as conclusões obtidas através da execução de vários experimentos e análise dos resultados.

TRABALHOS RELACIONADOS

A detecção de eventos no mar tem sido de grande interesse nos últimos anos. Instituições como a Agência Europeia de Segurança Marítima (EMSA)¹, Agência Espacial Europeia (ESA)² e a Administração Oceânica e Atmosférica Nacional (NOAA)³, têm-se dedicado a supervisionar eventos que acontecem no mar fornecendo relatórios periodicamente sobre eventos que representam um potencial risco para o meio ambiente.

O processo de estudo de eventos no mar pode incluir 3 focos possíveis, classificação ao nível da imagem, segmentação e detecção de objetos. A diferença entre essas abordagens será baseada em necessidades muito específicas, limitando cada uma a mostrar um rótulo por imagem (classificação), apresentar aos pixels um rótulo os quais contém a informação de interesse(segmentação) ou mostrando uma caixa em uma imagem apresentando a área de interesse.

Várias abordagens têm sido propostas para o estudo de eventos no mar: tendo abordagens baseados em algoritmos de aprendizado de máquina [7, 9, 10, 11, 12] (e.g. *Support Vector Machines* - SVM, *Random Forest*, Clustering) e aprendizado profundo [13, 14, 15, 16, 17, 18, 19] (e.g. Convolutional Neural Networks - CNNs). A identificação dos eventos pode ser centralizada em 2 focos, classificação ao nível da imagem e segmentação semântica (classificação definindo uma etiqueta por pixel). A utilização de cada enfoque será baseada na disponibilidade de dados e no tipo de análise do cenário de estudo.

O uso de métodos clássicos ligados a SVM, *Random Forest*, Clustering entre outros, tem se destacado por sua principal dificuldade em automatizar processos relacionados à geração de um sistema de classificação. Dados que descrevem uma cena de interesse, neste cenário dados vinculados ao Radar de Abertura Sintética, geralmente requerem uma extração especializada de atributos, exigindo um trabalho bastante significativo no estudo dos dados. Os dados relacionados ao Radar terá um fator de diferença com relação ao objetivo final destes, onde algumas configurações do sistema de medição retornarão uma maior ou menor quantidade de informações da cena de estudo.

¹<https://epso.europa.eu/>

²http://www.esa.int/Applications/Observing_the_Earth/Envisat/

³<https://www.ospo.noaa.gov/Products/ocean/marinepollution/>

Métodos baseados em *Random forest* têm mostrado um bom desempenho em tarefas de segmentação, onde a cena pode descrever mais de uma classe presente, especialmente com classes que tem valores de retroespalhamento semelhantes. Tong et al. [7] implementa uma abordagem baseada em *Random forest* para segmentação de manchas de óleo usando atributos polarimétricos. A premissa é focada na classificação supervisionada ao nível do pixel de um conjunto de dados com múltiplas polarizações e diferenças no ângulo de incidência. Na extração de atributos foi estudada a energia retroespalhada, correlação e estado polarizado entre diferentes canais, isto com o objetivo de extrair representações mais relevantes. Na inferência feita por *Random forest*, foi observado um bom desempenho em termos da métrica F1-score, tendo valores de 90,54% e *Precision* de 92,99%, usando dados SAR com valores entre (29.7° - 43.5°) no ângulo incidência.

Klapp et al. [9] mostra uma abordagem focada no uso de *Random forest* para a segmentação de áreas de gelo no mar. Apresenta um estudo de informações do Sentinel-1 para detecção de áreas com gelo marinho, destacando métodos vinculados à extração de atributos com base em matrizes de co-ocorrência em níveis de cinza (GLCMs). O abordagem mostra um foco nos valores de retroespalhamento para cada polarização disponível, otimizando a implementação do método *Random forest*. Conseqüentemente, é observado um resultados com *Precision* geral de 94,9%. Como conclusão é apresentada 2 questões importantes, como diminuir o alto gasto computacional e melhorar o desempenho em áreas de gelo marinho com presença de ruído, principais dificuldades do método proposto. Este tipo de métodos mostrou altos custos computacionais e um alto nível de análise para tratar problemas relacionados à segmentação de imagens de Radar [7, 9].

Chen et al. [10] propõe uma abordagem baseado em SVM para o estudo de imagens polarimétricas SAR. Os parâmetros de dissimilaridade, correlação, coerência, desvio padrão, grau de polarização, entropia, entre outros (GLCMs), são estudados tendo foco no estudo do efeito da polarização. A SVM apresentou pouca dependência em variações do tamanho dos dados (aprendizado robusto com imagens de tamanhos diferentes).

Baseado nos algoritmos de aprendizado de máquina, Joseph et al. [11] propõe uma abordagem para segmentação de imagens SAR usando SVM, realizando uma extração de atributos usando múltiplas polarizações. Neste cenário SVM mostrou um desempenho bom no estudo de imagens com variação no ângulo de incidência, tendo em contrapartida a necessidade de definir um critério de contexto e extração de atributos. Destaca-se uma análise profunda da extração de atributos, onde são estudados o atributo de textura,

isto avaliando a variância, homogeneidade, entropia, correlação, entre outros. Desta forma, através de uma análise visual, observa-se como SVM atinge bons resultados para segmentação supervisionada usando atributos de textura.

Xu et al. [12] apresenta uma abordagem baseado em *Local binary patterns* (LBP) e K-means (clustering), para a segmentação de derramamento de óleo. Neste trabalho é apresentada uma análise baseado na matriz de co-ocorrência de níveis de de cinza, tendo uma representação das variações no espalhamento. O modelo baseado em K-means demonstrou alto desempenho em tarefas de segmentação, sendo robusto a variações no ângulo de incidência nas imagens SAR. Por outro lado, a abordagem mostrou um pré-processamento baseado no uso de vários limiares para a representação do retroespalhamento, sendo necessária uma análise profunda para cada cena de interesse.

Métodos baseados em aprendizado profundo têm mostrado menor dependência de um especialista, isto por meio de uma extração aprendida de atributos, usando camadas convolucionais, tendo foco no aprendizagem de representações significativas. Durante seu treinamento é necessário ter uma grande quantidade de dados rotulados para aprender os parâmetros do modelo. Ao longo do tempo, as CNNs passaram por diversos avanços com o objetivo de melhorar o processo de aprendizagem de atributos e ao nível da otimização do custo computacional.

Al-Ruzouq et al. [13] faz uma revisão comparativa das abordagens baseadas em aprendizado de máquina para segmentação de eventos no mar. A partir do estudo apresentado, como contribuição mostra-se como escolher uma abordagem relacionada à classificação ou segmentação, com base no tipo de informação disponível. Como conclusão observa-se como a CNN consegue ter um desempenho melhor que o SVM, tendo um bom desempenho para detecção de óleo no mar (segmentação). Também é mostrado como a implementação de métodos baseados em CNN pode exigir uma grande quantidade de dados, os quais podem ter um custo alto devido aos processos de rotulagem.

A implementação de uma CNN está relacionada ao tipo de análise realizada no conjunto de dados alvo, ampliando ou reduzindo o nível de complexidade desse processo. Su et al. [14] fornece uma análise de um conjunto de dados aproveitando alguns conceitos relacionados à polarização de Radar e relacionados à CNN's, fazendo uso da arquitetura LeNet. Uma abordagem é apresentada a partir dos conceitos básicos da CNN, implementando uma arquitetura não tão profunda. O objetivo foi demonstrar as vantagens e desvantagens de otimizar o processo de obtenção de um modelo por meio da análise de atributos. Como vantagem, observa-se um alto desempenho para a classificação de objetivos no uso das polarizações HH (89,57% na *Precision*) e

VV (88,70% na *Precision*). Como conclusão, pode-se observar que esta análise é de vital importância, tendo diferentes níveis de importância para cada banda de polarização dependendo do cenário.

Yaohuaand Xudong et al. [15] enfatiza uma abordagem baseada em CNN's para segmentação de derramamento de óleo. O conjunto de dados empregado mostra eventos presentes no Mar da China, enfatizando uma extração multiescala de atributos. Ressalta-se como o uso de uma arquitetura baseada em DenseNet conseguiu alcançar bons resultados em um cenário com alto impacto de ventos fortes, onde os parâmetros de textura e intensidade tiveram uma variação significativa. Foi apresentada uma limitação relacionada ao tamanho do derramamento de óleo, mostrando um resultado de segmentação desfavorável para imagens com derramamento de óleo com poucos pixels em sua representação. Para os demais cenários, foi evidenciado um *Precision* na faixa de 95.35% - 96.77%.

O uso de dados de radar em conjunto com CNN tem se destacado por ter uma abordagem diferente aos métodos clássicos. Li et al. [17] mostra uma extração de atributos a partir do uso de camadas convolucionais com foco no atributo textura. Através da implementação do conceito ResNet, uma arquitetura ótima para classificação é implementada tendo valores médios na acurácia de até 89,65%. Como reflexão, é possível observar como a implementação da arquitetura baseada em ResNet atinge um bom desempenho na classificação de imagens de radar.

As CNN's têm mostrado uma ampla variedade de aplicações relacionadas no estudo de eventos no mar. Wang et al. [18] estuda diferentes eventos presentes no mar utilizando um modelo empírico na interpretação do retroespalhamento medido (CMOD5.N). Neste cenário foi proposto um modelo baseado na arquitetura Inception em conjunto com um modelo para interpretação de dados SAR, mostrando visualmente como CMOD5.N consegue gerar uma saída que descreve a cena corretamente. Porém, este trabalho mostra uma clara dependência do número de amostras por classe, havendo baixos valores de F1-score em classes com menor quantidade de imagens e menor variabilidade nas representações (0.64 na classe Ondas puras do oceano, 0.27 para *Iceberg*, e 0.56 para *Atmospheric front*).

O uso das CNN's na área de imagens SAR pode abranger um grande número de conceitos orientados ao uso de dados. Isso pode ser evidenciado ao estudar a banda de frequência utilizada na obtenção dos dados (ver [16, 17]) e ao fazer uma análise baseado na polarização (ver [14]). Da mesma forma, as CNN'S propõem várias perspectivas baseadas em necessidades bem definidas, por exemplo a extração esparsa de atributos (extração profunda de

atributos diminuindo o custo computacional, ver [18]) ou uso dos módulos de atenção(para o estudo de imagens com um tamanho grande ver [19]). As abordagens para o estudo de eventos no mar podem envolver uma visão de classificação ou segmentação. Com base na disponibilidade de dados, devido à dificuldade de rotulagem, abordagens voltadas para a classificação têm sido mais estudadas. Em conclusão as CNN's apresenta uma automatização no estudo de vários cenários com diferentes tipos de informação, sem ter a grande intervenção de um especialista. Como consequência, pode ser possível ter um modelo de classificação de alto desempenho que pode aprender um grande número de representações importantes.

Com base nisso, a dissertação terá como foco o estudo de métodos baseados em CNN, com o objetivo de demonstrar seu correto desempenho na classificação de eventos no mar, utilizando dados de radar, destacando o processo de automação para a obtenção de um modelo de classificação.

3 FUNDAMENTOS

Neste capítulo são expostos fundamentos relacionados ao sensoriamento remoto. Primeiro, é apresentada uma introdução ao sensoriamento remoto, mostrando tipos de sensores (ativos e passivos) e características do mapeamento, tendo uma visão geral do radar e do radar de abertura sintética (SAR). A seguir, serão apresentados aspectos relacionados ao SAR voltados ao mapeamento de áreas marinhas.

3.1 Sensoriamento Remoto

Considerando Lillesand et al. [20] o sensoriamento remoto pode ser definido como a ciência de obter informações sobre algum objeto de interesse, área ou evento por meio de um dispositivo que não está em contato físico com o objeto. Dependendo do princípio de funcionamento do sensor de interesse, é possível ter diferentes tipos de descrições da superfície de estudo, isso baseado em resolução, frequência e escopo. Entre esses sensores estão os espectrômetros, interferômetros, sensores baseados em radiometria, entre outros, que podem ser classificados como passivos ou ativos dependendo de seu princípio de funcionamento.

Atualmente, alguns sistemas usados em dispositivos como drones, aviões e satélites, estão sendo mais conhecidos, devido à seu uso em diversos setores da economia, como agricultura, monitoramento de áreas marinhas e estudo de áreas urbanas. Desta forma, o sensoriamento remoto tem sido implementado como uma solução eficaz para o estudo de grandes áreas, sendo de vital interesse para empresas e particulares. A Tabela 3.1 apresenta uma visão geral das plataformas utilizadas no ao sensoriamento remoto descrevendo as principais características de cada um.

Tabela 3.1: Plataformas usadas no sensoriamento remoto.

Plataforma	Características
Drone	<ul style="list-style-type: none"> * Baixo custo * Fácil acesso * Pequena extensão de mapeamento * Imagens com alta resolução
Satélite	<ul style="list-style-type: none"> * Grande área de mapeamento * Possui vários tipos de resolução espacial * Viabilidade de custo, envolve menor custo
Aeronaves	<ul style="list-style-type: none"> * Alto custo logístico * Pode ser viável em projetos de pesquisa de pequenas áreas * Resolução espacial média e alta

Sensores passivos e ativos

Sensor passivo na área de sensoriamento remoto refere-se a um dispositivo sem fonte própria de energia, que mede as energias refletidas pelos alvos. Sua atuação se concentra na medição da radiação refletida na superfície terrestre aproveitando possíveis fontes que interagem na área (fontes naturais ou artificiais externas). Alguns exemplos de sensoriamento remoto usando sistemas de sensores passivos podem ser relacionados ao uso de fotografia de filme, infravermelho e radiômetros. Alguns instrumentos usados são os Espectrômetros, Espectrorradiômetro, Radiômetro hiper espectral, Radiômetros de imagem e Sondas, os quais se baseiam no aproveitamento de fontes naturais de energia para a medição de características físicas de um objeto.

Por outro lado, um sensor ativo se concentra na emissão e recepção da radiação em uma superfície. São usadas fontes artificiais as quais funcionam em diferentes frequências, operando independentemente das condições de iluminação. Esses tipos de sensores são usados por um grande número de aplicações relacionadas à metrologia e observação da superfície terrestre. Dentre os instrumentos baseados em sensores ativos, destacam-se sistemas baseados em Radar, Lidar, Altímetro Laser e Scatterômetro.

3.2 Radar

3.2.1 Radar de Abertura Real

Os sistemas de detecção de radar são baseados na medição do eco. Milhares de pulsos em uma determinada frequência e polarização são emitidos, isto em torno da frequência de micro-ondas, para posteriormente medir o retorno do pulso emitido. Um radar típico (detecção de rádio e alcance) mede o índice de retroespalhamento da superfície de estudo e o tempo de retorno entre a emissão e recepção dos pulsos. Os sistemas baseados em radar usam pulsos onde a energia da antena é confinada em um curto intervalo de tempo, aquela energia estará relacionada às características físicas da antena. Por este motivo, os sistemas ligados à abertura real do radar serão limitados pelas características físicas da antena, conforme é mostrado na Figura 3.1.

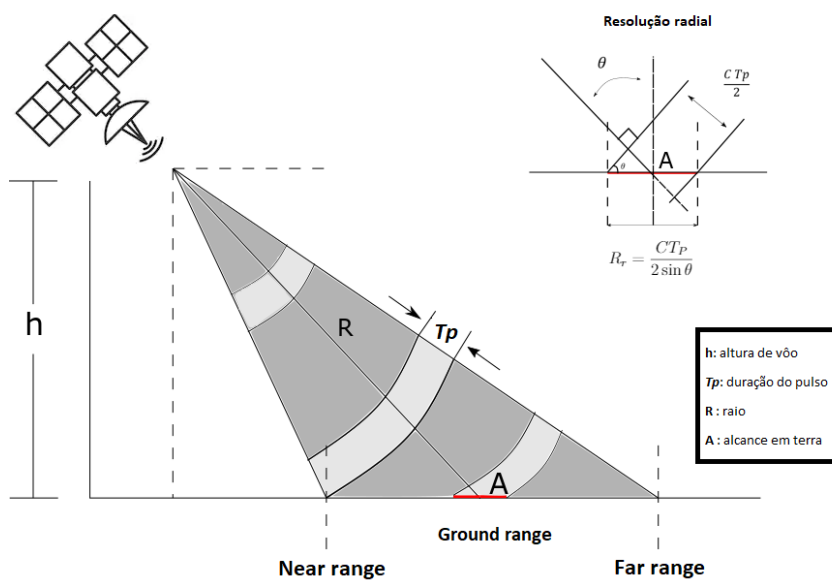


Figura 3.1: Resolução radial no processo de medição da radiação do espectro eletromagnético utilizando Radar.

Durante a abertura da antena, será medida uma área do terreno de interesse, definindo uma resolução no terreno, radial e de azimute. Com base nisso, a resolução radial R_r pode ser definida pela Equação 3-1, descrevendo a distância entre os centros de dois ecos pontuais, de tal forma que estes ecos não se sobreponham.

$$R_r = \frac{CT_P}{2 \sin \theta} \tag{3-1}$$

onde C representa a velocidade da luz, θ o ângulo de incidência, e T_p a duração do pulso.

No processo de mapeamento, a resolução de azimute no satélite estabelece um alcance na detecção, a Figura 3.2 mostra uma perspectiva superior que define a faixa da resolução de azimute.

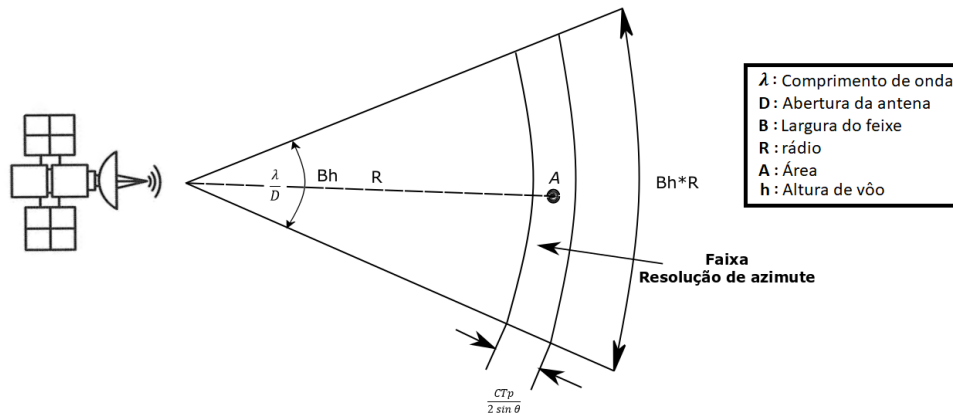


Figura 3.2: Resolução azimute no processo de medição da radiação do espectro eletromagnético utilizando Radar.

Nesse caso, o rango de azimute no para o radar de abertura real pode ser definida pela Equação 3-2.

$$B \cdot h \cdot R = \frac{\lambda h}{l \cos \theta} \quad (3-2)$$

onde h é a altura do vôo, B largura do feixe, R o rádio, λ o comprimento de onda e l o comprimento físico da antena.

A partir disso, o processo de emissão e recepção pode ser descrito por meio da Equação 3-3 chamada de Equação do Radar, que descreve o processo relacionado à abertura de Radar [20].

$$P_r = \frac{P_t G \lambda^2}{(4\pi)^3 R^4} \frac{\lambda R}{l} R_r \sigma_0 \quad (3-3)$$

onde P_r é o pulso refletido, P_t representa a potência transmitida, G é o ganho da antena, λ representa o comprimento da onda, R o alcance ao objeto-distância da antena receptora, R_r o rango do solo e σ_0 o índice de retroespalhamento.

Durante o estudo de uma área usando SAR, podem ser consideradas algumas técnicas as quais tem uma maior quantidade de informações da área de estudo, uma maior informação de altura (altimetria), representação das variações do terreno (interferometria) e uma melhor representação na amplitude do terreno (polimetria).

3.2.2 Radar de Abertura Sintética

Devido as limitações do Radar de Abertura Real (Real Aperture Radar - RAR), foi desenvolvido o Radar de Abertura Sintética (Synthetic Aperture Radar - SAR). No SAR é considerado um array de aberturas que simulam uma abertura maior mediante o movimento da antena, mapeando um areá maior sem necessidade de acrescentar o tamanho da mesma (ver Figura 3.3).

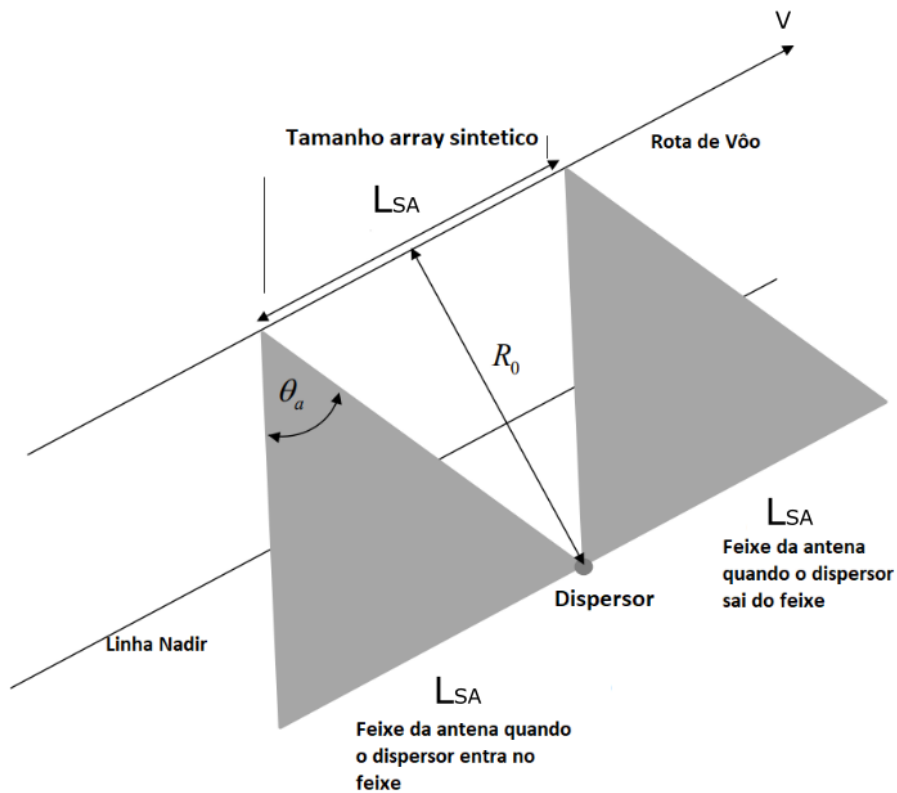


Figura 3.3: Processo de sensoramento usando radar de abertura sintética, obtido de [21].

O mapeamento SAR usa uma perspectiva baseada na geometria da superfície, medido pelo índice de retroespalhamento, o qual depende dos parâmetros do sistema e do terreno como é detalhado na Tabela 3.2.

Tabela 3.2: Parâmetros de aquisição no uso de radar e descrição do terreno de estudo.

Parâmetros do Sistema	Parâmetros do terreno
Polarização	Rugosidade
freqüência	Características dielétricas do terreno
Ângulo de incidência	Resposta da superfície

No caso do SAR, a resolução radial é definida pela Equação 3-1. A resolução do azimute é independente do intervalo R e é proporcional ao comprimento da antena, tendo grande influência no efeito doppler, durante as aberturas, levando em consideração a dinâmica na direção do vôo. O comprimento durante a abertura sintética (L_{SA}) pode ser definido pela Equação 3-4:

$$L_{SA} = \frac{\lambda R}{D_{AT}} \quad (3-4)$$

onde D_{AT} é o comprimento da antena ao longo do vôo e λ o comprimento de onda.

Para calcular a resolução de azimute do alvo, é necessário considerar uma mudança no efeito Doppler, razão pela qual durante as aberturas múltiplas elas se sobreporão, exigindo um cálculo de compressão para um rango de azimute. A Equação 3-6 mostra a compressão desta faixa, baseada em 2 rangos da antena, quando entra na faixa e quando sai da faixa do dispersor.

$$\delta_{AT} = \frac{\lambda}{2L_{SA}} R \quad (3-5)$$

Reduzindo a equação com base em L_{SA} .

$$\delta_{AT} = \frac{D_{AT}}{2} \quad (3-6)$$

Devido ao movimento na trajetória, há uma diferença na distância entre o radar e o espalhamento. Essa distância variável pode ser escrita com a Equação 3-7.

$$R(s) = \sqrt{R_o^2 + v^2 s^2} \quad (3-7)$$

onde R_o é o alcance mais próximo do dispersor, v a velocidade da plataforma do radar e s o tempo ao longo da trajetória de vôo.

Uma boa aproximação para radares de sensoriamento remoto é assumir que $vs \ll R_o$ (princípio de Radar). A partir de isso podemos aproximar o rango em função do tempo lento como na Equação 3-8.

$$R(s) \approx R_o + \frac{v^2}{2R_o} s^2 \quad (3-8)$$

A fase do sinal também mudará isso devido à compressão da faixa durante o cálculo de espalhamento. A partir disso, a fase é redefinida conforme mostrado na Equação 3-9. Como pode ser visto na Figura 3.2, a distância entre o radar e o espalhador muda, isso porque o espalhador pode estar diretamente ao lado da plataforma do radar. A mudança da distância entre o radar e o dispersor $R(s)$ significa que após a compressão do alcance, a fase do sinal será diferente para as diferentes posições ao longo da trajetória de voo.

$$\phi(s) = -\frac{4\pi R(s)}{\lambda} \tag{3-9}$$

3.2.3

Características dos sistemas SAR

No processo de mapeamento usando SAR alguns parâmetros devem ser considerados para a coleta de informações, isso porque dependendo desses parâmetros as informações conterão certas características de interesse. A Tabela 3.2 apresenta parâmetros relacionados ao sistema de radar e características presentes no terreno, os quais serão descritos nesta seção

Polarização

Durante a abertura do SAR, o pulso emitido terá uma polarização no campo eletromagnético, o que permitirá que o mapeamento seja representado de uma forma ou de outra. Este pulso pode ser polarizado em um campo vertical (V) ou horizontal (H), possuindo sistemas onde o pulso é emitido verticalmente e recebido horizontalmente (VH), ou emitido verticalmente e recebido verticalmente (VV), entre outras configurações (ver Figura 3.4).

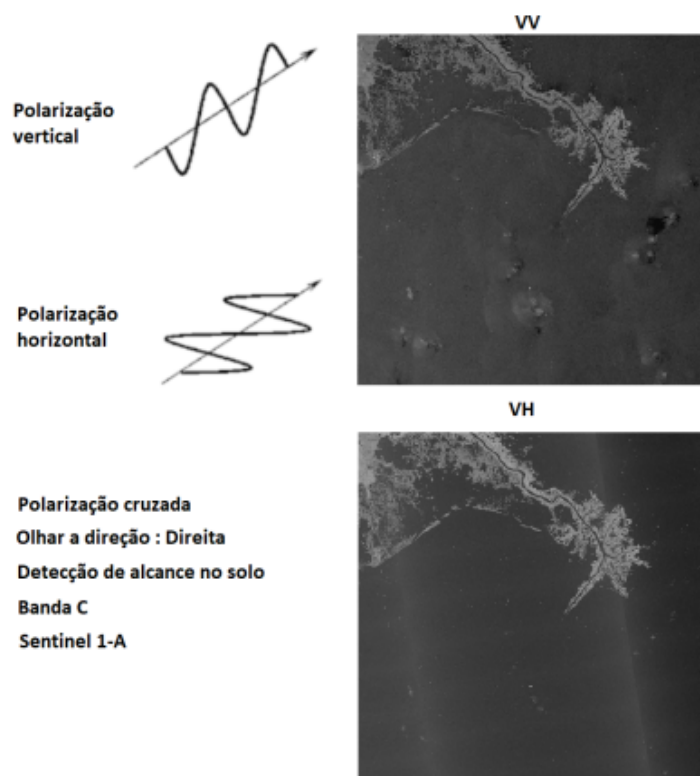


Figura 3.4: Polarização VH vs polarização VV para detecção de derramamento de óleo, produto de Sentinel-1A Level 1.

É comum que em algumas técnicas de mapeamento vários tipos de configurações sejam utilizados, obtendo-se diferentes tipos de informações (polarizações múltiplas). Dependendo da polarização de emissão e recepção, as informações capturadas podem ter uma melhor descrição no solo (VV) ou em altura (VH), isso devido à interação inerente ao campo eletromagnético.

Frequência

Trabalhar em uma determinada faixa de frequência permitirá obter informações mais detalhadas sobre a área de estudo. A Tabela 3.3 mostra as faixas de frequência presentes em cada banda.

Tabela 3.3: Bandas de frequência para SAR.

Banda	Comprimento de onda	Frequência (MHz)
Ka	0.75-1.15	40.000-26.500
K	1.1-1.67	26.500-18.000
Ku	1.67-2.4	18.000-12.500
X	2.4-3.75	12.500-8.000
C	3.75-7.5	8.000-4.000
S	7.5-15	4.000-2.000
L	15-30	2.000-1.000
P	30-100	1.000-300

Como mostra a Figura 3.5, dependendo da banda em que o radar atua, as imagens conterão maior ou menor informação sobre a área de estudo.

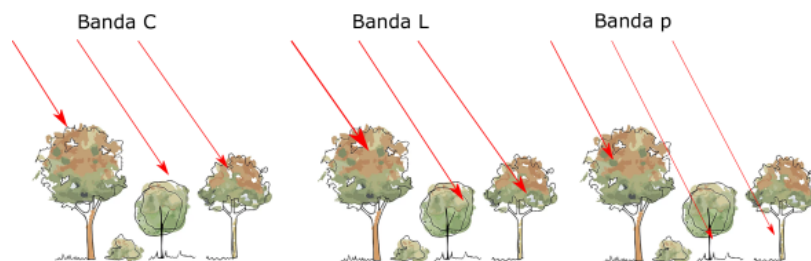


Figura 3.5: Efeito das bandas de frequência no mapeamento usando Radar.

Ângulo de incidência

Durante o mapeamento, é considerado os ângulos de incidência do pulsos de radar na superfície. Dependendo disso algumas áreas podem ser representadas escuras devido ao baixo retroespalhamento, em áreas com pouca incidência no pulso emitido, ou brancas tendo um retroespalhamento alto. Isso é mais evidente em áreas montanhosas, onde a zona atrás do objeto pode ser representada mais escura. A Figura 3.6 mostra como por não haver incidência do pulso emitido terá uma representação mais escura (encurtamento).

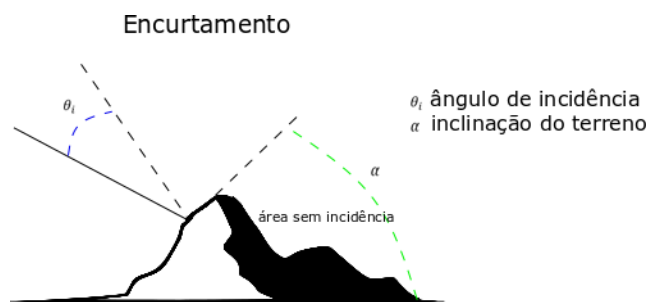


Figura 3.6: Processo de detecção baseado no ângulo de incidência, mostrando a geração de áreas com encurtamento.

Os processos de inversão e encurtamento estão relacionados ao estudo das distorções no mapeamento SAR. É comum que, devido às condições do terreno, o retroespalhamento tenha comportamentos diferentes destacando a presença de áreas mais escuras ou mais brilhante. A inversão de uma área no mapeamento é orientada a ter um baixo ângulo de incidência, onde o tempo de espera do espalhamento é maior em relação às demais áreas do terreno. Neste cenário é necessário aplicar um efeito de inversão, onde a área é representada como uma zona mais brilhante. Por outro lado, o encurtamento está relacionado a áreas escuras, como apresentada a Figura 3.7. Este efeito é baseado na diferença de altura na medição do terreno, onde a ponta do relevo fica mais próxima do sensor SAR. Com base nisso, as áreas mais distantes do sensor serão consideradas com menor espalhamento, devido ao princípio de funcionamento do Radar.

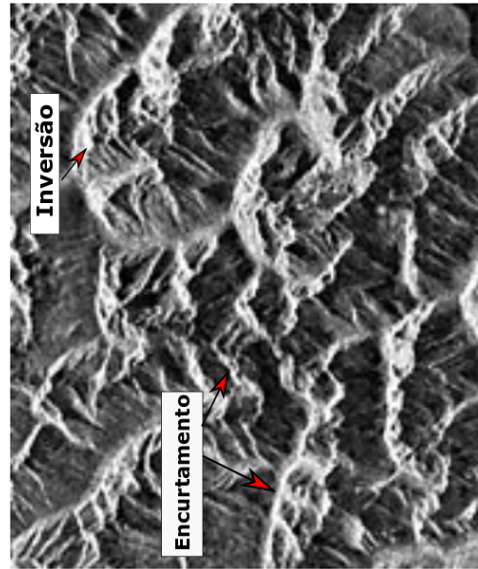


Figura 3.7: Distorções na imagem SAR: inversão e encurtamento. Obtido de [20]

PUC-Rio - Certificação Digital Nº 1912831/CA

Rugosidade

Esta característica depende do retroespalhamento do pulso emitido baseado nas características do terreno, como é mostrado na Figura 3.8. Dependendo da rugosidade da superfície, o retroespalhamento será diferente, o qual afetará diretamente o mapeamento da área.

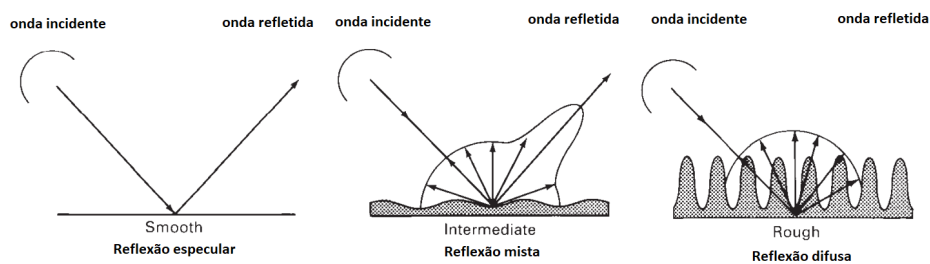


Figura 3.8: Espalhamento do pulso emitido com base nos tipos de superfícies, obtido de [20].

Este parâmetro intervém na representação de vários tipos de terreno, em termos gerais pode representar visualmente uma diferença entre áreas planas, com florestas, montanhas, cidade, áreas de cultivo, entre outras. Do

ponto de vista físico, o componente de incidência pode ter uma dispersão diferente dependendo das características do terreno. A Figura 3.9 mostra algumas representações comuns durante o mapeamento de radar, destacando uma diferença clara na rugosidade para cada área.

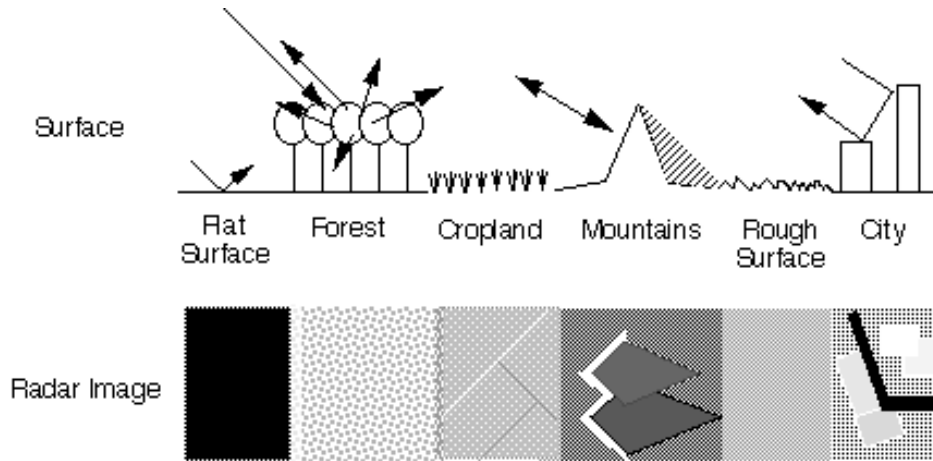


Figura 3.9: Alteração na imagem do radar a partir da variação da rugosidade da área de estudo, obtido de [20].

Características elétricas do terreno

Durante o sensoriamento por radar, o objetivo é estudar o retroespalhamento da superfície enviando um pulso durante a abertura. Nesse sentido, a permissividade dielétrica da superfície terá grande influência no mapeamento. A Equação 3-10 mostra a interação entre permissividade elétrica e índice de retroespalhamento.

$$\sigma_0 = \frac{4\pi R^2 |E_r|^2}{A_0 |E_c|^2} \quad (3-10)$$

$$E_r = \frac{E_c}{E_o} \quad (3-11)$$

onde, A_0 é a área de incidência, R é o rango mais próximo do dispersor, E_c a constante dielétrica do ambiente, σ_0 é o índice de retroespalhamento, E_r estará ligado à permissividade dielétrica relativa, baseado na permissividade do ambiente e a constante dielétrica no espaço livre $\frac{E_c}{E_o}$. A permissividade elétrica pode gerar representações mais claras ou mais escuras durante o mapeamento.

Resposta da superfície

Ao fazer o sensoriamento por radar, as características da área de interesse afetarão diretamente na representação final na imagem, principalmente devido ao índice de espalhamento. Essas representações descrevem a resposta do solo, Vegetação, Água e Gelo e Área Urbana (ver [20]).

Em cenários como o Mar é comum a presença de algumas classes relacionadas a eventos climáticos e atmosféricos. O mar está constantemente exposto a eventos meteorológicos, como fortes correntes de vento, chuvas, tempestades, baixas temperaturas, altas temperaturas, entre outros. Alguns eventos comuns são as frentes oceânicas (comportamentos físicos do mar), frentes atmosféricas (comportamentos causados por fatores climáticos), gelo, icebergs, áreas com pouco vento, manchas biológicas (manchas geradas por processos naturais), ondas do mar e células de chuva (vibrações geradas pela chuva) (ver [6]).

Os eventos no mar são caracterizados por ter uma representação distinta de rugosidade e intensidade. As zonas de vento fraco podem ser representadas como áreas homogêneas mais escuras, as manchas biológicas com representações escuras em forma de uma linha ou espiral, células de chuva com linhas contínuas representando rugosidade, frentes atmosféricas com uma linha escura que limita uma área, onde há uma mudança na intensidade.

3.3

Radar para monitoramento de áreas marinhas

A detecção de Radar pode cobrir um grande número de cenários, incluindo áreas marinhas. Para o cenário relacionado ao mar, teremos uma menor presença de terrenos de grande altitude devido à presença em maior extensão de água. Com base nisso, polarizações centralizadas na obtenção de informações da superfície serão de maior interesse, apresentando maior importância nas polarizações VV, tendo mais informações para o eixo horizontal. A configuração VV tem se destacado principalmente em uma série de aplicações, incluindo o estudo da rugosidade em pequena escala para ondas de superfície no mar.

Informações das bandas C e L

Durante o processo de mapeamento do radar, diversos parâmetros relacionados à configuração do sistema de monitoramento de uma área estão envolvidos. Com base nisso, as informações do estudo variariam significativa-

mente, tendo diferentes níveis de representação. No caso de áreas marinhas, essas representações irão variar no que diz respeito à resolução e quantidade de informações de um possível evento de interesse. Um exemplo disso pode ser o estudo das camadas de gelo marinho, onde será possível ter uma detecção do evento de superfície (basta detectar a presença de gelo) ou uma descrição detalhada do evento (espessura da camadas de gelo para classificação).

Para o estudo de áreas marinhas, algumas bandas (frequência) de interesse são definidas para trabalhar durante o mapeamento por radar. O efeito que terá no resultado dos dados do estudo estará relacionado à obtenção de uma diferença na resolução espacial (representação de uma área em um pixel) e na quantidade de informação na área de estudo (retroespalhamento). A Tabela 3.3 apresenta as bandas comumente usadas para monitorar áreas marinhas.

Dentre as bandas utilizadas em radar para monitoramento de áreas marinhas, é mais comum ter informações de imagem utilizando a banda L e a banda C. A banda L é caracterizada por conter uma maior quantidade de informações da área marinha e é comumente utilizada para o estudo do gelo no mar e Iceberg. Em função disso, a banda L permitirá estudar eventos mais específicos como a variação das camadas de gelo após um processo de mudança atmosférica, mostrando possíveis deformações presentes nessas camadas de gelo. Por outro lado, a banda C oferece uma resolução menor do que a banda L, sendo suficiente para detectar a presença de um evento, como a presença da camada de gelo. Em contraste o uso da banda L pode ter mais informações das mudanças específicas dentro desses eventos (detalhes da cena), dependendo do uso da banda C ou L, a imagem do radar que representará uma área pode ter uma descrição correta do evento. (ver Figura 3.10).

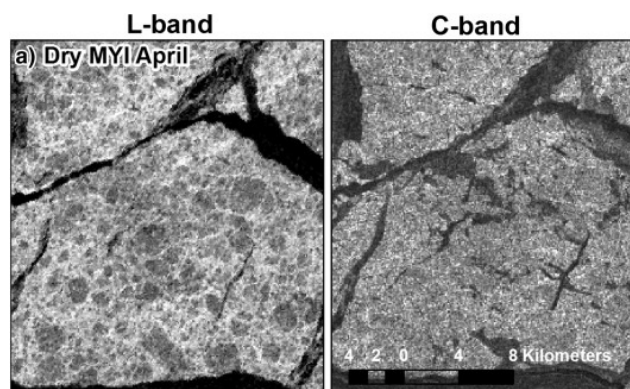


Figura 3.10: Banda C vs Banda L mapeamento no mar (Gelo marinho), obtido de [22].

Em relação à polarização, conforme indicado nas seções anteriores, a

polarização VV orientada para uma perspectiva horizontal retornará uma quantidade maior de informações no mar. A Figura 3.4 mostra o efeito que a polarização tem para o estudo de uma mesma área no mar.

4

REDES NEURAI CONVOLUCIONAIS PARA CLASSIFICAÇÃO

Esta seção apresenta os fundamentos teóricos para as arquiteturas VGG16, Inception v3, Xception, Inception-ResNet v2, MobileNet e módulos baseados em atenção.

4.1

Conceitos básicos

As redes neurais convolucionais (CNN) são um tipo de rede neural inspirada na visão humana. Neste, são considerados arrays de grande dimensão, que podem corresponder a séries temporais expressas em arrays 1D, imagens em arrays 2D (pixels), arrays de dados cúbicos 3D, entre outros. Compreendem uma série de camadas compostas por neurônios onde a partir de uma entrada (I) e um peso (w) para as conexões entre os neurônios, pode gerar uma saída expressa em pontuações para cada classe. Essas camadas de neurônios serão conectadas entre elas, sendo a última camada totalmente conectada expressamente ao número de classes de interesse. Como no ser humano, tenta-se simular as sinapses entre neurônios, onde um estímulo maior pode gerar um efeito positivo ou negativo no processo de aprendizagem.

Em [23], LeNet-5 apresenta uma das primeiras abordagens orientadas para o reconhecimento de dígitos, onde técnicas relacionadas ao aprendizado baseado em gradiente e retro-propagação de gradiente são estudadas e expostas, apresentando uma aplicação importante para o uso de dados representados em 2D, expresso em imagens de dígitos. Com base nessa contribuição, foram propostas arquiteturas cada vez mais profundas com uma maior quantidade de camadas. Um dos aspectos mais importantes destacados pelo LeNet-5 é a presença de *vanishing gradient*, marcando uma limitação clara em profundidade na arquitetura proposta. O efeito de *vanishing gradient* mostra um fenômeno presente durante o processo de aprendizagem, onde ter uma rede mais profunda pode gerar que a derivação do gradiente convergir para um valor que não é definido como um mínimo global, gerando que o processo de aprendizagem não continue. A Figura 4.1 mostra a arquitetura LeNet-5, onde a entrada foi definida em uma dimensionalidade $32 \times 32 \times 1$, seguido por 2 pares de camadas

de convolução com *stride 2* e camadas de *average pooling* com *stride 1*.

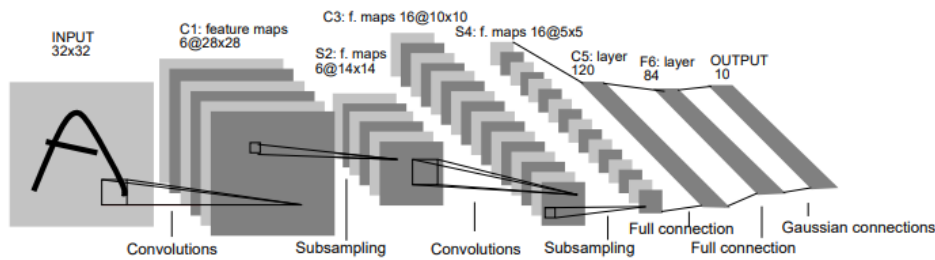


Figura 4.1: LeNet-5. obtido de [23].

Tendo em conta que um dos objetivos traçados na implementação da CNN é o aprofundamento evitando o *vanishing gradient*, AlexNet [24] apresenta uma abordagem robusta onde é superada a pontuação anterior para o desafio no conjunto de dados ImageNet. Da contribuição destaca técnicas relacionadas ao aumento de dados para reduzir o sobre ajuste e a inclusão de não linearidades com a camada de ativação ReLU, reduzindo o efeito do *vanishing gradient*.

Conforme mencionado acima, a inclusão de camadas convolucionais tem ajudado no sucesso da CNN, devido à sua versatilidade e ótimo desempenho no aprendizado de atributos relevantes, permitindo menor dependência de um especialista para extrair atributos.

4.1.1 Camadas convolucionais

As camadas convolucionais extraem atributos, definindo um *kernel* e aplicando a operação matemática de convolução usando um *stride*. A Figura 4.2 mostra um exemplo referente à operação de Convolução, definindo uma imagem de entrada de tamanho 6×6 e um *kernel* de 3×3 .

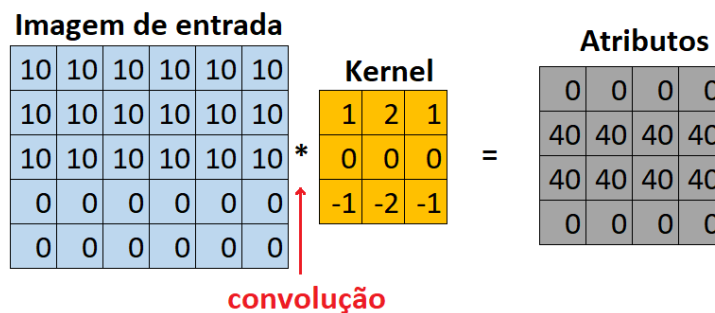


Figura 4.2: Exemplo da operação de Convolução.

A camada de convolução pode ser definida pela Equação 4-2.

$$A_{out}[M, N] = (I_{array} * K)[M, N] \quad (4-1)$$

$$(I_{array} * K)[M, N] = \sum_j \sum_h k[j, h]f[M - j, N - h] \quad (4-2)$$

onde A_{out} são os atributos de saída, I_{array} a imagem de entrada, k o *kernel*, M e N as coordenadas em um array 2D e j, h o deslocamento nas 2 dimensões.

4.1.2

Camada de Pooling

As camadas de pooling fazem uma redução dos atributos mais representativos com base no critério de máximo (*Max pooling*) ou média (*Average Pooling*). Para a redução, é considerado um *kernel* (geralmente 2×2) que será usado para calcular um máximo ou média.

A Figura 4.3 mostra um exemplo relacionado à operação max pooling. Neste caso, um *kernel* e uma *stride* são definidos, e a partir deste os valores serão agrupados com base no valor máximo presente na região atual definida pelo *kernel*.

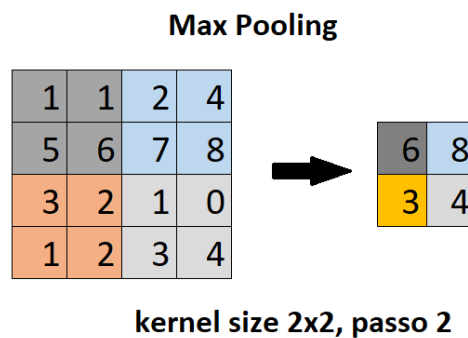


Figura 4.3: Operação de Max Pooling: imagem 4x4, *kernel* 2x2, *stride* = 2.

4.1.3

Camada de Dropout

Um dos efeitos negativos durante o treinamento da CNN é o *overfitting*, que impede um bom aprendizado. A camada dropout descarta certas conexões aleatoriamente, isto baseado em um porcentagem. A Figura 4.4 mostra uma representação de 50% de dropout para uma camada intermediária.

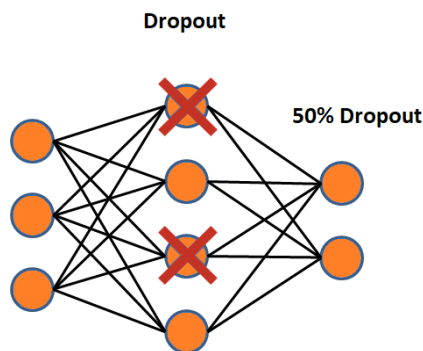


Figura 4.4: Operação de Dropout: 50% de Dropout.

4.1.4

Camada totalmente conectada

Esta camada permite a conexão dos neurônios presentes nas camadas anteriores, levando em consideração as ativações anteriores. Como resultado, o cálculo é reduzido a uma multiplicação de uma matriz, que representa as conexões anteriores.

4.2

Métricas de avaliação

4.2.1

Matriz de confusão

A matriz de confusão descreve o desempenho de um modelo de classificação. O objetivo é representar a taxa de verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN). O objetivo no estudo da matriz é ter a maior taxa de verdadeiros positivos para cada classe.

		Positive	Negative
Classe prevista	Positive	TP	FP
	Negative	FN	TN

Figura 4.5: Matriz de confusão: cálculo dos verdadeiros positivos(TP), verdadeiros negativos(TN), falsos positivos(FP) e falsos negativos(FN) para uma classe prevista.

A Figura 4.5 apresenta uma matriz de confusão, onde as áreas verdes definirão o comportamento desejado na classificação (classificação correta de positivos e negativos).

4.2.2

Precision

A métrica Precision apresenta a razão entre os verdadeiros positivos e todos os positivos, o objetivo é ver o quão correta é a inferência de uma amostra. A Equação 4-3 mostra como calcular o *Precision*.

$$Precision = \frac{TP}{TP + FP} \quad (4-3)$$

onde *TP* (*True positive*) são as amostras da classe de interesse classificadas corretamente, e *FP* (*False positive*) amostras que não pertencem à classe e são classificadas incorretamente.

4.2.3

Recall

Um dos aspectos de interesse ao avaliar o desempenho de um modelo é analisar a quantidade de verdadeiros positivos classificados corretamente. A Equação 4-4 mostra como o Recall é calculado.

$$Recall = \frac{TP}{TP + FN} \quad (4-4)$$

onde *FN* (*False negative*) são amostras da classe classificadas incorretamente.

4.2.4

F1-score

A métrica de F1-score é definida como a média harmônica entre as métricas *Precision* e *Recall*, assim como na Equação 4-5.

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4-5)$$

4.2.5

GradCAM

Durante o processo de avaliação de um modelo de classificação, é possível estudar os mapas de ativação presentes para a definição de uma classe. Idealmente, esta área deve conter informações relevantes sobre o evento, objeto ou cenário de estudo.

O GradCAM apresenta uma análise baseada na geração de mapas de ativação presentes durante a classificação de uma amostra.

$$\alpha_k^c = \frac{1}{z} \sum_i \sum_j \frac{\delta y^c}{\delta A_{ij}^k} \quad (4-6)$$

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (4-7)$$

A Equação 4-7 apresenta o cálculo dos pesos de importância do neurônio α_k^c , descrevendo a linearização parcial da rede profunda ajustada para A (ativações), capturando um nível da importância dos atributos k , isto tendo em conta a classe de interesse c . Obter o mapa de localização discriminativo de classe GradCAM, envolverá considerar a combinação linear do produto da matriz de peso, onde a ativação ReLU é aplicada ([25]).

4.3

Arquiteturas estudadas

4.3.1

VGG16

Apresentado pela universidade de Oxford em [26], a VGG16 apresenta uma arquitetura que teve um alto desempenho no desafio ILSVRC-2014, obtendo um Precision nas imagens de teste de 92,7% no conjunto de dados ImageNet. Em comparação, VGG16 melhora o desempenho do AlexNet ao substituir os *kernel* de grande tamanho por vários *kernel* de tamanho 3×3 um após o outro. A entrada é definida com um tamanho de $224 \times 224 \times 3$, que passará por um conjunto de camadas seguidas por um *kernel* de tamanho de 3×3 e após uma convolução 1×1 . O *pooling* espacial é realizado por cinco camadas de *Max-Pooling*, que seguem algumas das camadas convolucionais. O *Max-Pooling* é realizado em uma janela de 2×2 pixels, com *stride* 2. Posteriormente, três camadas totalmente Conectadas seguem as pilhas de camadas convolucionais, onde as 2 primeiras têm 4096 canais cada uma, a terceira tem foco na classificação e portanto, contém 1000 canais (um para cada classe) e a camada final é a camada softmax. A Figura 4.8 mostra a arquitetura exposta acima.

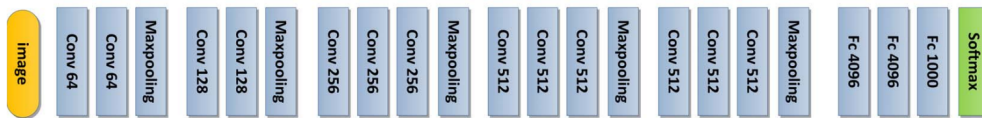


Figura 4.6: Modulo de VGG16, obtido de [26].

4.3.2

Inception V3 (Incep)

A arquitetura Inception V3 [27] apresentada no ano 2015 pela universidade College London, expande o conceito do Inception, melhorando as pontuações impostas por suas versões anteriores no desafio ILSVRC 2012. O Inception V3 apresenta em sua implementação o conceito de divisão em convoluções menores, tendo convoluções assimétricas, classificadores auxiliares baseados no uso de *Batch Normalization* e redução do tamanho do *Grid*, consolidado em uma mesma arquitetura. O objetivo é permitir uma melhor eficiência computacional reduzindo o número de parâmetros (23.83 milhões) envolvidos na rede para ter um treinamento mais rápido.

Em sua implementação, o Inception V3 apresenta a fatoração de convolução como forma de reduzir o número de parâmetros presentes na arquitetura sem afetar seu desempenho. A Figura 4.7 mostra o exposto acima, onde se busca uma fatoração em pequenas convoluções, onde um *kernel* 5×5 tem 25 parâmetros mas pode ser fatorado em duas convoluções com *kernel* de tamanho 3×3 . As pequenas convoluções tem apenas 18 parâmetros ($3 \times 3 + 3 \times 3$).

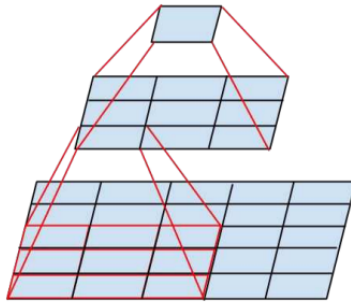


Figura 4.7: Fatoração de convoluções 5×5 em convoluções menores 3×3 , obtido de [27].

Além das convoluções na fatoração, o Inception V3 estuda como manter um bom campo receptivo aproveitando o processo de fatoração.

A Figura 4.8 mostra como as convoluções originalmente consideradas no bloco de Inception são modificadas em convoluções. As convoluções assimétricas são aplicadas, onde uma convolução 3×3 pode ser substituída por uma convolução 1×3 seguida por uma convolução 3×1 , tendo um número de parâmetros um pouco menor do que a convolução original. Neste cenário o objetivo é fazer uma redução de *Grid*, isto para ter um foco em representações de alta dimensão.

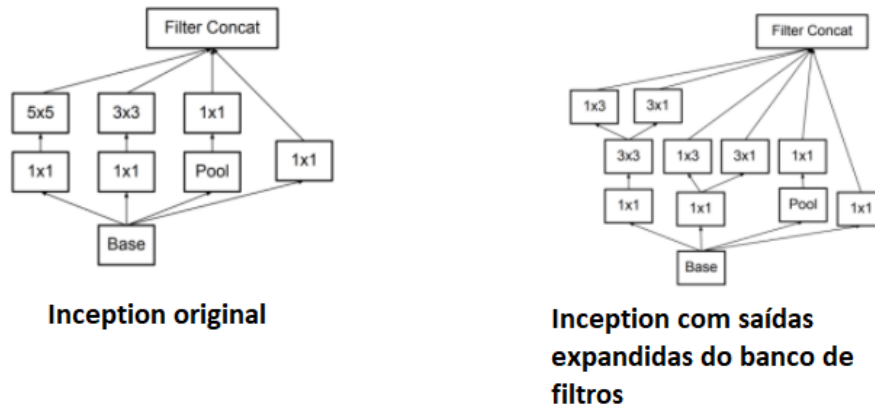


Figura 4.8: Inception original vs módulo de Inception proposto ($Grid\ 8 \times 8$), obtido de [27].

Para alcançar uma maior profundidade, o Inception V3 propõe o uso de um classificador auxiliar. Conforme mostrado na Figura 4.9 o classificador auxiliar é inserido entre as camadas durante o treinamento tendo um efeito regularizador, isto ajuda a ter um melhor desempenho na extração de atributos de baixo nível.

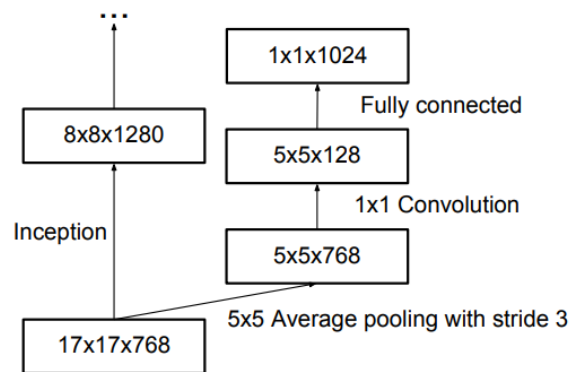


Figura 4.9: Implementação do classificador auxiliar como regularizador: O Classificador auxiliar é usado na parte superior da última camada 17×17 , tendo um Batch Normalization, ganho Top-1 error 0,4% , obtido de [27].

4.3.3 Xception (Xcep)

Baseado no Inception, o Xception [28] propõe blocos extremos de Inception. Xception aproveita o conceito de convoluções separáveis (convoluções normais no mesmo nível), onde a operação de convolução é dividida em con-

voluções menores localizadas no mesmo nível. Isto ajuda a ter uma rede computacionalmente mais rápida. No Xception, 36 camadas convolucionais são estruturadas em 14 módulos, todas com conexões residuais lineares ao seu redor, exceto o primeiro e o último módulo (tendo 22.8 milhões de parâmetros).

Em resumo, a arquitetura Xception poder ser definida como uma pilha linear de camadas de convolução separáveis em profundidade com conexões residuais. Como mostra a Figura 4.10, a arquitetura primeiro usa uma convolução 1×1 para mapear correlações entre canais mapeando separadamente as correlações espaciais de cada canal de saída. Xception em resumo apresenta uma convolução separável em profundidade modificada, onde apenas uma convolução 1×1 será considerada na entrada, isto sem considerar uma linearidade ReLU intermediária. Isso permite maior diversidade no que diz respeito à extração de atributos e atinge maior profundidade e baixo custo computacional, devido ao efeito de regularização presente no módulo Inception.

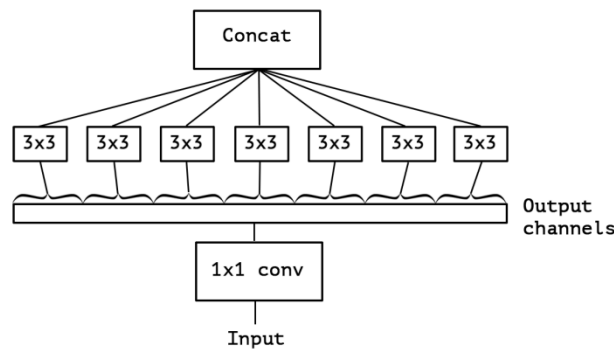


Figura 4.10: Módulo de Xception, obtido de [28].

Devido ao seu funcionamento, os conceitos apresentados na Xception, relacionam uma extração de atributos considerando um campo receptivo, usando conexões residuais para obter uma maior profundidade e tendo um foco na separação baseada na profundidade, algo semelhante a *depthwise separable convolution*, em especial pelo uso extremo do Inception. Os 2 conceitos podem ser diferenciado, *depthwise separable convolution* realiza primeiro a convolução espacial em canais e, em seguida, realizam a convolução 1×1 , enquanto o mostrado no Xception é feito uma convolução 1×1 primeiro.

Em [28] se apresenta como o Xception consegue melhores resultados que os modelos VGG16, ResNet e Inception V3 sobre a base de dados ImageNet (Tabela 4.1).

Tabela 4.1: Top-1 e Top-5 da métrica Precision no conjunto de dados ImageNet, obtido de [28].

Modelo	Top-1 accuracy	Top-5 accuracy
VGG16	0.715	0.901
ResNet-152	0.770	0.933
Inception V3	0.782	0.941
Xception	0.790	0.945

4.3.4 Inception ResNet v2 (Incep-R)

Inception ResNet apresenta o uso dos conceitos de conexão residual e modulo de Inception, isso com o objetivo de ter uma rede mais profunda. O objetivo do ResNet Inception é propor uma arquitetura híbrida, onde o conceito de conexão residual será incluído na saída da operação de convolução presente no módulo Inception para a entrada do modulo. A arquitetura está focada em evitar a fuga do gradiente (*vanishing gradient*, morte do gradiente, onde o mínimo global não pode ser alcançado.), ao ter uma rede tão profunda.

Como é mostrado na Figura 4.11, cada bloco de Inception é seguido por uma camada de convolução 1×1 sem ativação, que é usada para aumentar a dimensionalidade dos atributos antes da adição, isto para coincidir com a profundidade da entrada.

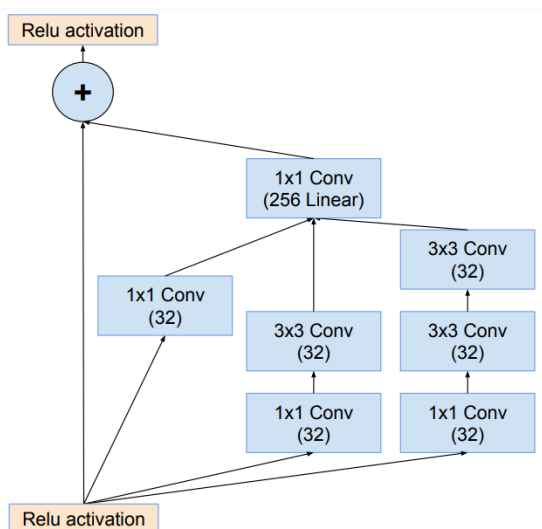


Figura 4.11: Bloco Inception com conexões residuais, obtido de [29].

Experimentalmente foi demonstrado que ao usar mais de 1000 *kernels*, a aprendizagem é afetada, isso por causa de que a última camada antes do agrupamento médio produz apenas zeros após de algumas dezenas de milhares de iterações.

A partir disso, um dos grandes desafios e contribuições que a Inception ResNet proporciona é a necessidade de aplicar técnicas para atingir maior profundidade, de forma que a rede se torne estável. Geralmente, adicionar camadas de normalização ou reduzir a taxa de aprendizagem ajuda a evitar a estagnação no treinamento. Observando que isso não é experimentalmente suficiente para o Inception ResNet, esta arquitetura propõe uma redução dos resíduos para controlar o efeito das conexões residuais.

Os resíduos da saída do bloco de Inception serão usados, após de uma camada de convolução, isto baseado em uma regularização dos resíduos. A regularização está relacionada a ter uma maior estabilidade na rede, onde o objetivo é definir uma escala nas ativações residuais. Experimentalmente foi mostrado em [29], que considerar este processo ajudou a ter uma maior estabilidade para uma profundidade maior que 1000 *kernel*. No bloco de *Activation scaling* mostrado na Figura 4.12, fatores de escala entre 0,1 e 0,3 são aplicados. No bloco Inception ResNet, convoluções com *kernels* de vários tamanhos são combinados com conexões residuais, isto para diminuir o tempo de treinamento do modelo.

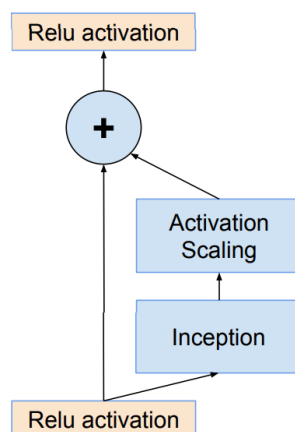


Figura 4.12: Dimensionamento dos resíduos usando uma ativação com escalonamento, obtido de [29].

O modelo foi avaliado no conjunto de dados ILSVRC 2012, onde é possível observar um melhor desempenho em relação às demais arquiteturas baseadas em Inception (ver Tabela 4.2).

Tabela 4.2: Top-1 e Top-5 do erro para as arquiteturas baseadas em Inception no desafio ILSVRC 2012, obtido de [29].

Modelo	Top-1 erro	Top-5 erro
Inception V3	21.2%	5.6%
Inception V4	20.0%	5.0%
Inception-ResNet V1	21.3%	5.5%
Inception-ResNet V2	19.9%	4.9%

4.3.5 MobileNet (MNet)

A arquitetura MobileNet tem como foco apresentar uma estrutura com poucos parâmetros e baixa complexidade. Em [30], a MobileNet apresenta uma arquitetura baseada no uso dos conceitos de convolução separável em profundidade (ver Figura 4.13), multiplicador de largura e multiplicador de resolução.

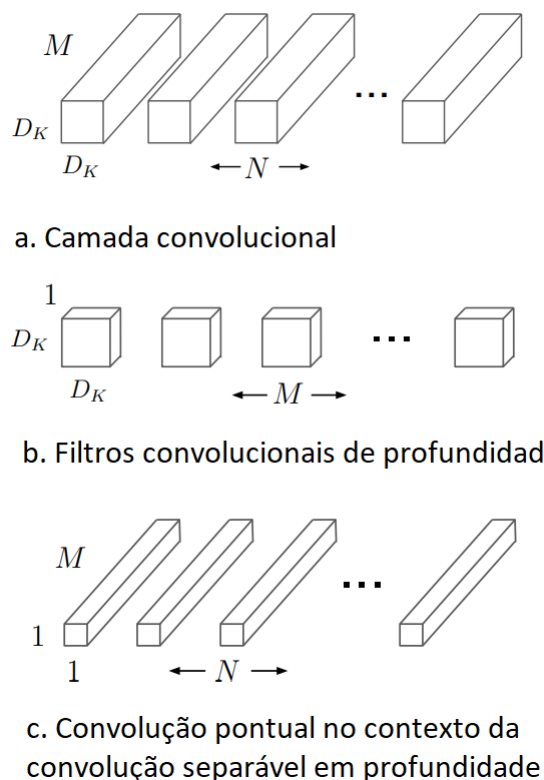


Figura 4.13: Convolução separável em profundidade, obtido de [30].

A convolução separável em profundidade surge da ideia de separar

uma convolução normal com o objetivo de reduzir o custo computacional. Como apresenta a Figura 4.13, é possível separar uma convolução maior em outras menores, para posteriormente considerá-las usando uma camada convolucional 1×1 . A camada convolucional normal é parametrizada pelo *kernel* de convolução K de tamanho $D_K \times D_K \times M \times N$, onde D_K é a dimensão espacial do *kernel* assumido como quadrado, M o número de canais de entrada e N o número de canais de saída. Os atributos de saída para a convolução normal, pode ser definido como é mostrado na Equação 4-8.

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \quad (4-8)$$

A Equação 4-9 mostra o custo computacional usando convoluções padrão, isto com base no número de canais de entrada M , do número de canais de saída N , o tamanho do *kernel* $D_K \times D_K$ e o tamanho dos atributos $D_F \times D_F$.

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \quad (4-9)$$

As convoluções separáveis em profundidade são usadas para quebrar a interação entre o número de canais de saída e o tamanho do *kernel*. A operação de convolução normal tem o efeito de filtrar atributos com base nos *kernels* convolucionais e combinar atributos para produzir uma nova representação. As etapas de filtragem e combinação podem ser divididas em duas etapas (*depthwise convolution* e *pointwise convolution*) por meio do uso de convoluções fatorizadas chamadas *depthwise separable convolutions*.

Depthwise convolution faz uso de um única convolução por cada canal de entrada (profundidade de entrada). *Pointwise convolution*, usa uma convolução 1×1 , o objetivo é criar uma combinação linear da saída da camada de profundidade. A Equação 4-10 mostra a saída em atributos, onde \widehat{K} é o *depthwise convolutional* são de tamanho $D_K \times D_K \times M$.

$$\widehat{G}_{k,l,m} = \sum_{i,j,m} \widehat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m} \quad (4-10)$$

A partir disso, uma grande redução no custo computacional é gerada, conforme mostrado pela Equação 4-11. Além disso, cada *Depthwise convolution* é seguida por uma camada de batch normalization e uma não linearidade, com o objetivo de alcançar uma maior profundidade. O downsampling é feito usando convolução com *stride* nas *Depthwise convolution*. Uma camada de *average pooling* reduz a resolução espacial antes da camada totalmente conectada.

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F \quad (4-11)$$

4.3.6 Deep Attention Sampling (DAS)

A arquitetura esta baseada no uso de *Attention Sampling*, onde é proposto a extração de atributos a uma escala menor baseado no nível de importância da amostra. A imagem de entrada passa por um módulo de geração de *patches* baseado em atenção, onde o objetivo será extrair um conjunto de *patches* menores da imagem (sub-imagens) baseado na estimativa de Monte Carlo, onde é definindo os atributos mais importantes presentes nos *patches* obtidos, baseado no uso da rede de atenção.

Conforme mostrado na Figura 4.14, a arquitetura apresentada terá como foco a geração de Q *patches* por meio de uma estimativa Monte Carlo, isto baseado no uso do módulo de atenção espacial. Como entrada a arquitetura terá 2 escalas de uma imagem, executando um processamento de *downsampling* para uma dessas imagens. Para a implementação da amostragem por atenção, será apresentada uma distribuição de sub-imagens que passarão pelo módulo de atenção. Como resultado, apenas alguns *patches* Q da imagem de entrada irão para a rede convolucional completa. Com base nisso, a arquitetura permitirá uma abordagem multiescala, onde a geração dos *patches* terá como foco detectar as regiões mais relevantes e gerar várias sub-imagens.

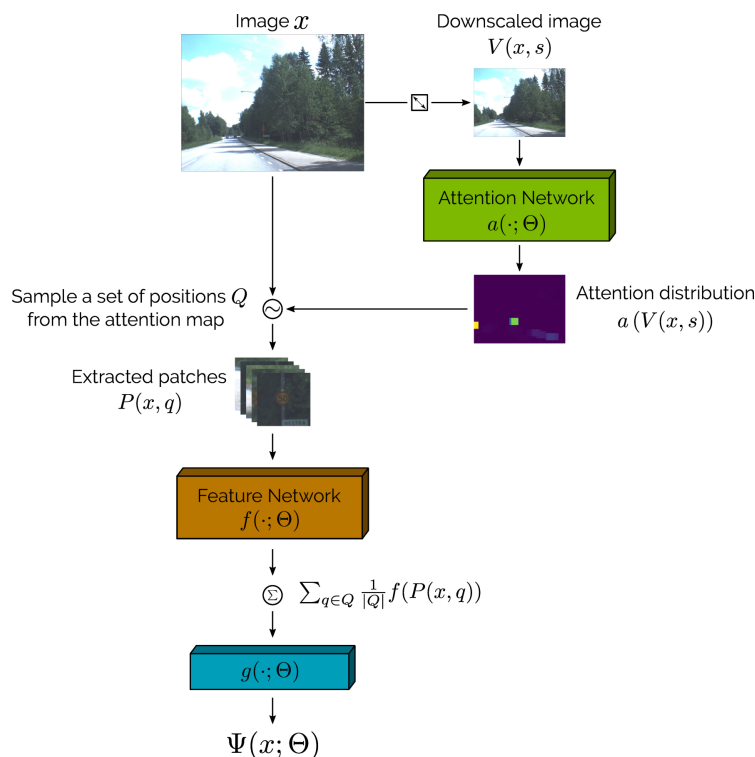


Figura 4.14: Amostragem por atenção usando uma rede totalmente convolucional, obtido de [19].

$$\psi(x; \Theta) = g\left(\sum_{i=1}^K a(x; \Theta)_i f(x; \Theta)_i\right) \quad (4-12)$$

A Equação 4-12 mostra como a amostragem de atenção intervém explicitamente na derivação do gradiente. Onde $A(\cdot)$ é o módulo de amostragem baseado em atenção, $f(\cdot)$ representa a *feature network* implementada e $g(\cdot)$ é a camada de classificação linear.

5 METODOLOGIA

Considerando o cenário proposto relacionado ao estudo de eventos no mar, foi usada uma metodologia voltada ao uso de CNN para classificação. A abordagem centrou-se no estudo de 2 conjuntos de dados relacionados com eventos no mar (com e sem pré-processamento). A Figura 5.1 apresenta as principais etapas consideradas na metodologia: Geração de *patch* para treinamento e teste, Treinamento dos modelos de classificação, Inferência com os *patch* de teste, e Construção do mapa de classificação.

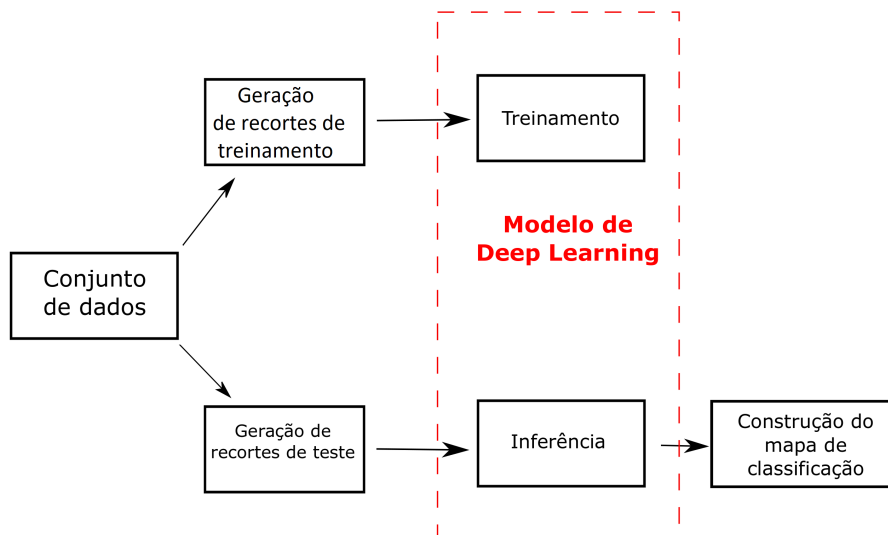


Figura 5.1: Metodologia para gerar os mapas de classificação usando modelos baseados em aprendizado profundo.

5.1 Pré-processamento dados Sentinel-1A

As imagens SAR do Sentinel-1 precisam passar por um pré-processamento para corrigir efeitos relacionados ao ruído térmico, diminuição do ruído Speckle, calibração radiométrica e transformação para dB. A Figura 5.2 mostra o grafo de processamento feito as imagens Sentinel-1 na ferramenta SNAP da *European Space Agency* (ESA) baseado no [32].

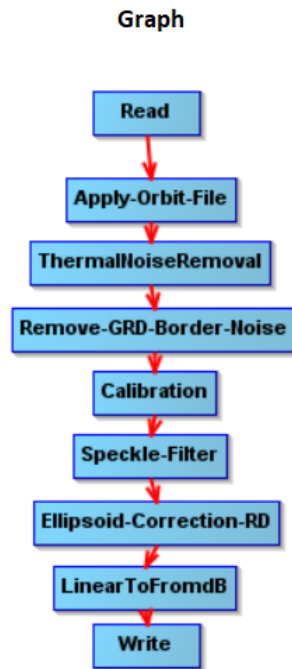


Figura 5.2: Grafo de pré-processamento para as imagens Sentinel-1 do conjunto de dados de derramamento de óleo.

Para a execução do processo de treinamento, foram considerados apenas os dados pré-processados. No caso do conjunto de dados de derramamento de óleo, o processamento proposto nesta seção foi implementado. O objetivo principal do pré-processamento estará focado em realizar operações de filtragem e correção radiométrica e de orbita. Desta forma, o arquivo de saída estará disponível em formato GeoTIFF geo-referenciado, possuindo 2 canais, um para polarização VV e outro para polarização HH .

5.2

Construção do conjunto de dados

A partir do conjunto de dados de imagens SAR com suas respectivas etiquetas, foi gerado dois conjuntos de dados mutuamente excludentes: treinamento e teste. No caso que as imagens SAR sejam de um tamanho maior ao tamanho de entrada dos modelos de classificação $w \times h$, é necessário fazer uma extração de *patch* do mesmo tamanho, $w \times h$, com determinado passo ou *stride*. Além disso, se tem alguns casos onde as etiquetas fornecidas são para Segmentação Semântica (a nível de pixel), onde é necessário gerar as etiquetas para a classificação ao nível de *patch*. Estas etiquetas a nível de *patch* são geradas considerando a proporção da classe de interesse presente no *patch* (ver

Equação 5-1). Se essa proporção fosse maior a um determinado limiar, o *patch* é considerado como parte da classe de interesse.

$$Proporção_{classe} = \frac{\#Pixels_{classe}}{\#Pixels_{total}} \quad (5-1)$$

onde $\#Pixels_{classe}$ e $\#Pixels_{total}$ representam o número de pixels da classe de interesse no *patch* e o total de pixels, respectivamente.

Definir um tamanho de *patch* vai influenciar o contexto das informações estudadas, onde um dado, dependendo da presença de informações, pode indicar uma tendência direta ou indireta para o mesmo evento.

A Figura 5.3 apresenta três exemplos de contexto alto, médio, e baixo, mostrando como a partir da quantidade de informações a descrição de uma imagem pode mudar. Nessa perspectiva, descrever corretamente um evento por meio de uma imagem pode mudar significativamente, apresentando grandes dificuldades em cenários onde não temos um bom tamanho de *patch*. Isso está relacionado a um possível evento precursor ao surgimento da classe de interesse.

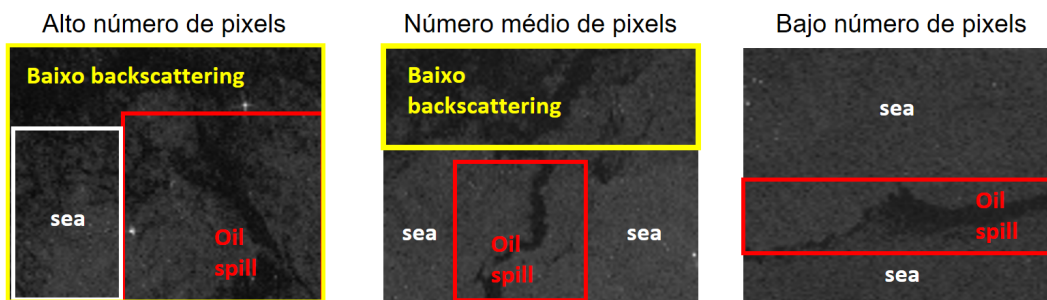


Figura 5.3: Contexto com base na quantidade de informações em um tamanho de *patch*.

Para definir o tamanho correto do *patch*, neste cenário foram realizados experimentos com os seguintes tamanhos de *patch* 128×128 , 256×256 , 380×380 e 512×512 .

5.3 Implementação de modelos baseados em CNN

Os modelos considerados apresentados no Capítulo 4: VGG16, Inception v3, Xception, Inception ResNet v2, MobileNet e o Deep Attention Sampling foram implementados. Nesta seção, foi estudado 6 modelos baseados em CNN, isso com o objetivo de delimitar uma área de estudo, devido à grande diversidade de arquiteturas baseadas em aprendizagem profunda hoje presentes. A seleção dessas 6 arquiteturas baseou-se nos focos de cada uma, tendo como

fator mais importante o custo computacional e o desempenho de classificação em cenários relacionados ao uso de dados de radar. Para a arquitetura Deep Attention Sampling foi implementado um módulo de atenção para amostragem e foi considerado a implementação de uma seção totalmente convolucional baseada no VGG16.

Definidas as arquiteturas de interesse, foi executado um treinamento baseado no uso de *patches* de cenas maiores. O ajuste dos parâmetros de treinamento baseou-se em um protocolo experimental, fazendo variações no otimizador, taxa de aprendizado e normalização, com base no estudo da métrica de f1-score do treinamento. O estudo desta métrica refletirá de que forma o modelo está aprendendo e se está caminhando para o caminho desejado.

Dentro da seção de estudo de arquiteturas baseadas em CNN, foi proposto estudar os efeitos relacionados às vantagens do uso de módulos de atenção para amostragem e várias arquiteturas focadas no uso de convoluções esparsas em profundidade.

5.3.1 Sintonização de hiperparâmetros

Um dos aspectos mais importantes para o processo de obtenção do modelo de classificação é o ajuste dos hiperparâmetros. Neste cenário, será possível sintonizar 2 aspectos, o primeiro relacionado à construção do conjunto de dados e o segundo relacionado aos parâmetros de treinamento e inferência.

Como foi mostrado na seção anterior, a construção do conjunto de dados estará relacionada aos fatores de contexto e relevância das informações de uma cena, definindo diferentes tamanhos de imagem e *stride* na geração dos *patches*.

No caso de ajuste dos parâmetros de treinamento, as métricas propostas (pontuações f1-score, *Precision* e *recall*) serão consideradas para validação durante o treinamento. Com base nisso, serão definidos o otimizador, a taxa de aprendizado e o número de épocas para cada treinamento. Para isso, ferramentas em forma de *callbacks* serão utilizadas, automatizando o processo exploratório de ajuste de parâmetros. Foi implementado um *early stopping* evitando sobre treino, uma taxa de aprendizagem adaptativa e um *model checkpoint*, evitando modelos com métricas inferiores às observadas durante todo o treinamento.

Na inferência, foram consideradas 2 análises, GradCAM e construção de um mapa de inferência. O tamanho do *patch* será relacionado ao modelo treinado e a passada durante a construção do mapa de inferência será fixada em 100%. Para a definição das camadas de ativação no GradCAM, ela será

avaliada a partir da última camada de *pooling* definida, utilizando o total de camadas superiores.

5.3.2

Processo de Inferência

O processo de inferência envolveu um estudo de cenas no nível do *patch*. Para avaliar o desempenho do modelo treinado no conjunto de teste se empregam métricas como o *Precision*, *Recall* e F1-score a nível de classe. Além disso, o GradCAM pode ser usado para avaliar quais foram as áreas de maior interesse para gerar a previsão do modelo.

Para a análise de resultados foi necessário fazer a construção do mapa de classificação juntando os *patches* do mesmo jeito que foram gerados, mantendo a distribuição espacial entre eles. Nesse cenário, a construção do mapa de inferência não estudo cenários com *patches* sobrepostos (Figura 5.4).

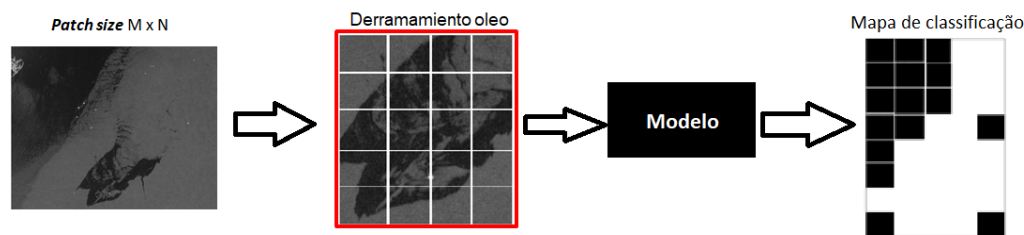


Figura 5.4: Processo de inferência usando *patches*.

Obter o mapa de inferência refletirá em termos gerais o desempenho correto do modelo de classificação. Considerando o interesse no estudo de grandes áreas, a obtenção de um resultado de inferência georreferenciada das mesmas dimensões da entrada (considerando o CSR) será ideal para a análise dos resultados.

O processo de inferência foi feito ao nível de cena para o conjunto de dados de Derramamento de Óleo no Mar, apresentando o mapa de classificação e GradCAM. Para o caso do conjunto de dados de eventos no mar, o GradCAM foi avaliado ao nível de *patch* em conjunto com as métricas de inferência, devido à disposição dos dados.

6 ANÁLISE EXPERIMENTAL

Esta seção apresenta aspectos relacionados aos experimentos realizados neste trabalho. Primeiro são descritos os conjunto de dados empregados nos experimentos. Em seguida são apresentadas as configurações dos parâmetros usadas nos modelos avaliados e o protocolo experimental. Finalmente, os resultados obtidos nos experimentos são apresentados e analisados.

6.1 Conjuntos de Dados

Para o estudo de eventos no mar foram considerados 2 conjuntos de dados com imagens SAR. O primeiro mostra um conjunto de dados o qual descreve vários cenários presentes no mar. O segundo conjunto de dados descreve um evento relacionado ao derramamento de óleo ocorrido próximo à costa do México.

6.1.1 Eventos Naturais no Mar

Este conjunto de dados é apresentado por Wang et al. [6] e corresponde a vários eventos no mar (ver Tabela 6.1). As imagens foram obtidas no 2016, como mostra a Figura 6.1, em uma grande extensão dos oceanos Antártico, Índico e Arábico. O conjunto de dados está composto por imagens SAR Sentinel-1A Wave Mode adquiridas em faixas de $20km$ de largura, com $5km$ de resolução e polarização VV. A Tabela 6.1 mostra as classes presentes no conjunto de dados assim como o número de imagens por classe.

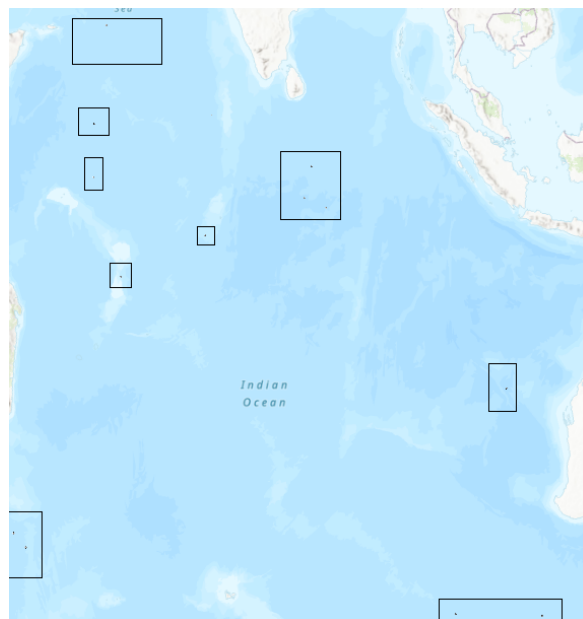


Figura 6.1: Localização da área de estudo para o conjunto de dados de Eventos Naturais no Mar.

Tabela 6.1: Distribuição do conjunto de imagens para Eventos Naturais no Mar, dados obtidos próximo ao oceano Antártico, Índico e Arábico no 2016.

Classe	#Imagens
Ondas puras do oceano	4900
Rajada de vento	4797
Células micro-convectivas	4598
Células de chuva	4738
Manchas biológicas	4709
Gelo marinho	4369
Icebergs	1980
Área de vento fraco	2159
Frente atmosférica	4097
Frente oceânica	1199

O conjunto de dados contém 10 classes (ver Figura 6.2). Ondas puras do oceano (Figura 6.2a) representa um movimento típico das ondas no mar. Rajada de vento (Figura 6.2b) descreve uma perturbação gerada por uma circulação contrária do vento, tendo presença de representações periódicas em forma de linha (superfície onde circula o vento para cima e para baixo). Células micro-convectivas (Figura 6.2c) descreve estruturas geradas a partir da diferença de

temperatura local entre o ar e o mar, gerando um ar ascendente/descendente em forma de célula. As Células de chuva apresentadas na Figura 6.2d descrevem as vibrações expressas na rugosidade gerada pelo impacto da chuva ao interagir com a superfície. As Manchas biológicas (Figura 6.2e) mostram alguns pontos relacionados a processos naturais, onde o retroespalhamento é menor (áreas escuras). O Gelo marinho (Figura 6.2f) descreve áreas de baixa temperatura, onde a superfície é congelada uniformemente, criando uma camada no mar, e Iceberg (Figura 6.2g) descreve pedaços maiores de gelo liberados em torno de uma área congelada maior. As Área de vento fraco (Figura 6.2h) destacam-se por apresentar baixa retrodifusão devido à pouca movimentação do mar, possuindo representações escuras que podem cobrir um perímetro significativo. As Frentes atmosféricas apresentadas na Figura 6.2i têm uma representação em forma de uma linha na borda da classe ou massa, esta gerada por variações ambientais expressas em temperatura ou vento, tendo representações de áreas de baixo e alto vento. As Frentes oceânicas mostradas na Figura 6.2j representam camadas que limitam duas massas de água diferentes, isso expresso em salinidade e densidade oceânica.

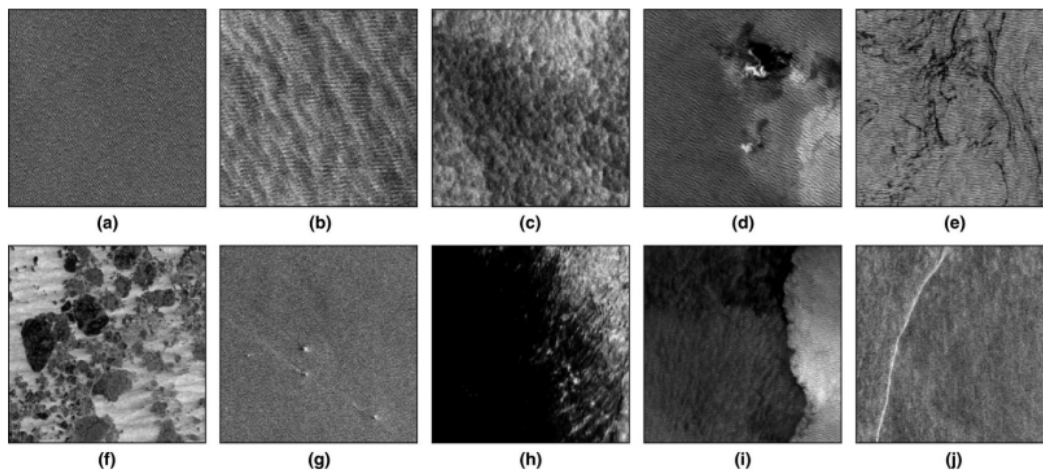


Figura 6.2: Imagens das classes presentes no conjunto de dados de Eventos Naturais no Mar, (a) Ondas puras do oceano, (b) Rajada de vento, (c) Células micro-convectivas, (d) Células de chuva, (e) Manchas biológicas, (f) Gelo marinho, (g) Icebergs, (h) Área de vento fraco, (i) Frente atmosférica, and (j) Frente oceânica, obtido de [6]

A polarização VV no radar da banda C descrever melhor a área de interesse no mar, principalmente no que diz respeito à sua rugosidade. Neste caso, a *sea surface roughness* (*ssr*) expressa uma representação dos dados do radar

baseada em um modelo de interpretação empírica, onde o retroespalhamento terá uma recalibração. A representação de ssr expressa as características da área de estudo, como foi estudado nas seções anteriores, os valores de retroespalhamento estão altamente relacionados aos parâmetros utilizados durante o mapeamento por radar. Neste caso, o modelo empírico CMOD5.N mostrado em [31] calcula a rugosidade da superfície com base nos parâmetros padrão do sistema, onde a constante de vento é $10 \frac{m}{s}$, assumindo um ângulo de abertura relativo de 45° , como mostrado na Equação 6-1.

$$ssr = \frac{\sigma_0}{CMOD5.N(10, 45, inc)} \quad (6-1)$$

O conjunto de dados disponibiliza a informação em 2 tipos de dados. O primeiro apresenta os dados no formato GeoTiff de 16bits, onde foi feito um processo de recalibração de retroespalhamento e re-amostragem, tendo uma representação em um intervalo entre [0; 65,535]. O segundo apresenta uma versão em PNG de 8 bits, onde as imagens passaram por um processo de recalibração, re-amostragem e quantização, obtendo uma representação em um intervalo entre [0-255]. Nos experimentos foram usadas as imagens PNG de 8 bits, seguindo o que é proposto em [18], com fins de comparar nossos resultados com um baseline.

6.1.2

Derramamento de Óleo no Mar

Este conjunto de dados foi formado por imagens SAR do derramamento de óleo no Golfo do México. Como pode ser visto na Figura 6.3, a área de interesse está localizada perto da costa da Louisiana. O conjunto de dados tem anotações feitas por especialistas do NOAA¹ em formato de vetores geo-referenciados², devido a isso foi necessário procurar as imagens SAR correspondentes a cada anotação. As imagens SAR são produtos nível 1 do Sentinel-1A com projeção de detecção de alcance de terra (GRD) e polarização VV, obtidos da Copernicus Open Access Hub³.

¹Administração Nacional Oceânica e Atmosférica

²<https://www.ospo.noaa.gov/Products/ocean/marinepollution/>

³<https://scihub.copernicus.eu/dhus/#/home>

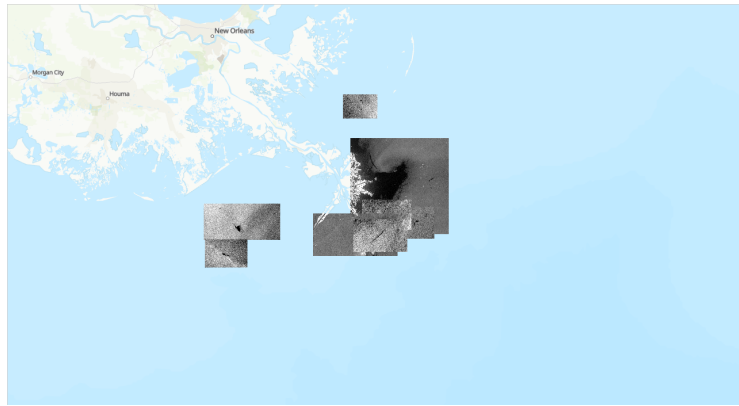


Figura 6.3: Área de interesse para o conjunto de dados de Derramamento de Óleo no Mar.

Os relatórios fornecidos pela NOAA contêm descrições de confiabilidade, junto com informações sobre o local e a data em que as imagens foram obtidas. Para a construção do conjunto de dados, foram consideradas só as anotações com alta confiabilidade (ver Figura 6.4).

PUC-Rio - Certificação Digital Nº 1912831/CA

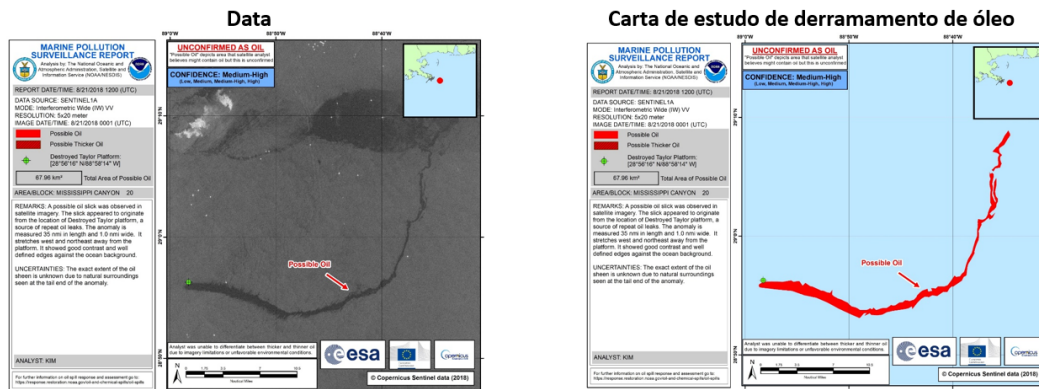


Figura 6.4: Conjunto de relatórios obtidos na NOAA para eventos no mar do Golfo do México. Os relatórios mostram o vetor da anotação junto com as coordenadas, órbita e data da imagem de radar estudada.

A Tabela 6.2 apresenta as imagens SAR no conjunto de dados, as respectivas datas de aquisição e o tamanho de cada imagem. Também é mostrado as imagens que foram usadas para o treinamento e teste dos modelos avaliados.

Tabela 6.2: Conjunto de dados de Radar para derramamento de óleo, obtidos entre 2018-2020. As imagens em negrito são usadas para treinamento e as outras para teste.

Data	Tamanho	Data	Tamanho
21/08/2018	5800×4600	16/08/2019	3900×2000
14/09/2018	3800×3200	08/09/2019	4100×1900
26/09/2018	5200×2900	15/10/2019	5100×2100
13/11/2018	4600×2600		4400×2200
07/12/2018	4900×2600	24/02/2020	4000×2100
	4900×2300	07/03/2020	3900×2400
	5300×2500	19/03/2020	3800×2100
	4700×2200		4100×2000
19/12/2018	2500×2000	31/03/2020	3700×2100
	3900×2400	22/08/2020	3600×2300
	3600×2600	-	-
31/12/2018	2900×3600	-	-
	5700×3000	-	-
		-	-

6.2

Configuração Experimental

Baseado na metodologia mencionada no Capítulo 5, foram empregadas 6 arquiteturas para classificação de eventos no mar: Inception, Xception, Inception ResNet, MobileNet VGG16 e Deep Attention Sampling usando os dois conjuntos de dados apresentados.

Os *patches* foram extraídos das imagens SAR no conjunto de dados de Derramamento de Óleo no Mar considerando dois tamanhos: 256×256 e 512×512 , e um *stride* de 10%.

A Tabela 6.3 e 6.4 mostra a quantidade de *patches* gerados. Para a geração das etiquetas foi considerada uma proporção de 0.05 (ver Equação 5-1).

Tabela 6.3: *Patches* gerados do conjunto de dados de Derramamento de óleo no Mar considerando um tamanho de 512×512 e *stride* 10%.

Classe	#Imagens
Óleo	4890
Não óleo	4520

Tabela 6.4: *Patches* gerados do conjunto de dados de Derramamento de óleo no Mar considerando um tamanho de 256×256 e *stride* 10%.

Classe	#Imagens
Óleo	10125
Não óleo	9420

A Tabela 6.5 apresenta as arquiteturas avaliadas nos experimentos. Foram usados os pesos pré-treinados com o conjunto de dados ImageNet para cada modelo. As camadas superiores foram removidas e substituídos por camadas densas principalmente e no final uma camada densa com número de neurônios igual ao número de classes do conjunto de dados.

Tabela 6.5: Arquiteturas avaliadas na classificação de eventos no mar.

Modelo	Top Layer	# Parâmetros
Xception(Xcep)	Top layer(layer-130, moving_mean)+GlobalAvgPooling2D + dense(2048, activation=relu) + dropout(0.5)+ dense(1024, activation=relu) + dropout(0.5)+ dense(512, activation=relu) + dropout(0.5)+ dense(N_class, activation='softmax')	Total = 27,681,834 Parâmetros Treináveis = 6,874,882
Inception(Incep)	Top layer(layer-310,mixed 7) + GlobalMaxPooling2D + dense(2048, activation=relu) + dropout(0.3)+ dense(1024, activation=relu) + dropout(0.3) + dense(512, activation=relu) + dropout(0.3) + dense(N_class, activation='softmax')	Total = 28,623,138 Parâmetros Treináveis = 6,820,354
Inception ResNet(Incep-R)	Top layer(layer-780,conv_7b_ac) + GlobalMaxPooling2D + dense(1024, activation=relu) + dropout(0.5) + dense(512, activation=relu) + dropout(0.5) + dense(N_class, activation='softmax')	Total = 57,486,050 Parâmetros Treináveis = 3,179,586
VGG16(VGG16)	Top layer(19,block5_pool) + dense(1024,activation=relu) + dense(512,activation=relu) +dropout(0.3) + dense(N_class, activation='softmax')	Total = 15,765,826 Parâmetros Treináveis = 2,887,170
MobileNet(MNet)	Top layer(layer-153,out_relu) + global_average_pooling2d + dense(1024,activation = relu) +dropout(0.5)+dense(512,activation = relu) + dense(N_class, activation='softmax')	Total = 5,145,154 Parâmetros Treináveis = 2,887,170
Deep Attention Sampling(DAS)	Attention sampling module (size = 512, scale = 0.2, patch_size =70x70, multinomial_entropy) + VGG16 Non top +GlobalMaxPooling+L2Normalize()+ Dense(classes,activation=softmax)	Total = 14,716,595 Parâmetros Treináveis = 1,907,00

A Tabela 6.6 apresenta os parâmetros considerados no treinamento dos modelos, definindo uma distribuição para o conjunto de dados de treinamento de 70%, 20% para validação e 10% para teste. Durante o processo de treinamento foi considerada uma etapa de normalização junto com um aumento de dados.

Tabela 6.6: Parâmetros de treinamento.

Parâmetro	Valor
Normalização	Padrão (média=0, desvio padrão=1)
Optimizador	Adam (momentum = 0.9)
Taxa de aprendizado	0.001
Épocas (max)	100
Distribuição	Train = 70%, Validation =20%, Test =10%
Early stopping	F1-score sem melhora em 6 épocas
Aumento de dados	Flip vertical, flip horizontal, Zoom

Na implementação do modelo Deep Attention Sampling foi usado um valor de $scale = 0.2$ e um $patch_size = 70 \times 70$. O valor de $scale$ define que outros tamanhos serão utilizados a partir da entrada, lembrando que a entrada passará por um processo de re-amostragem, apresentando ao módulo de atenção a escala original e a escala definida no parâmetro. O $patch_size$ define o tamanho das sub-imagens obtidas através da atenção.

6.3

Protocolo experimental

No processo de treinamento, foi considerado um total de 40 experimentos por arquitetura, tendo uma média entre 60-100 épocas. O processo de treinamento teve como foco o estudo da métrica F1-score, visando uma boa relação entre Precision e Recall. As abordagens foram focadas em um problema binário para o caso do conjunto de dados de Derramamento de Óleo no Mar e multi classe para o conjunto de dados de Eventos Naturais no Mar.

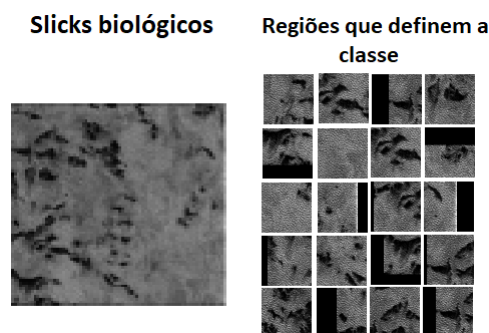


Figura 6.5: *Patches* gerados pelo módulo de atenção. Os *patches* foram geradas a partir da estimativa de Monte Carlo considerando uma atenção espacial, com base na derivação do gradiente.

Para o processo de treinamento baseado no Attention sampling, foi considerado uma atenção espacial para a extração de *patches*. A Figura 6.5 mostra um exemplo dos *patches* gerados para uma entrada, com foco nas áreas de maior interesse. Desta forma, a descida do gradiente será baseada na operação do módulo de atenção conforme é mostrado em [19].

O treinamento foi executado em um cluster com quatro GPUs NVIDIA Tesla V100 de arquitetura Volta com 16 GB de RAM e 5120 núcleos Cuda cada uma.

6.4 Resultados

6.4.1 Eventos Naturais no Mar

Avaliação do Treinamento

A Figura 6.6 apresenta a evolução do F1-score durante o treinamento dos modelos avaliados. Cada curva representa a média do F1-score ao longo das épocas de treinamento para as 40 rodadas de cada modelo.

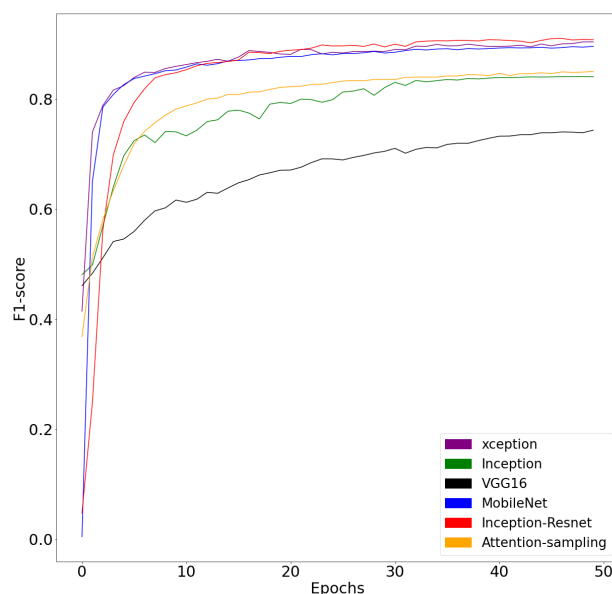


Figura 6.6: Curvas de treinamento no conjunto de dados de Eventos Naturais no Mar: F1-score vs. Número de Épocas. O gráfico apresenta a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.

Os modelos Inception ResNet, Xception e MobileNet atingiram os maiores valores de F1-score: 0.91, 0.90 e 0.89, respectivamente. Em seguida, Deep Attention Sampling e Inception obtiveram valores de F1-score de 0.85 e 0.84. A VGG16 conseguiu o menor valor de F1-score: 0.73.

Para complementar a análise dos valores de F1-score obtidos por cada modelo foi calculado o intervalo de confiança das 40 rodadas para cada modelo. A Figura 6.7 apresenta estes intervalos de confiança para cada modelo. O modelo mais estável em termos de intervalo de confiança é o Deep Attention Sampling atingindo valores com uma variação de ± 0.02 . Em relação à outras arquiteturas, Inception ResNet teve um valor de ± 0.03 , VGG16 de ± 0.03 , Xception de ± 0.04 , MobileNet de ± 0.05 e Inception de ± 0.06 .

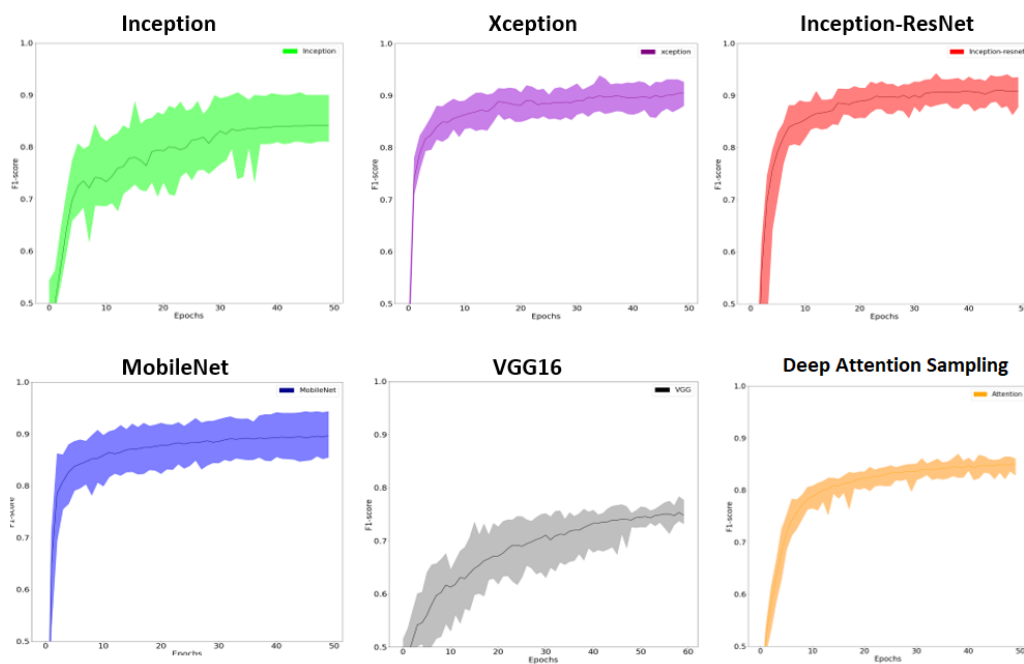


Figura 6.7: Intervalo de confiança das curvas de treinamento no conjunto de dados de Eventos Naturais no Mar: F1-score vs. Épocas. O gráfico apresenta o intervalo de confiança para a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.

Considerando o valor da métrica F1-score e estabilidade dos modelos durante o treinamento, o Inception ResNet teve o melhor desempenho obtendo o maior valor de F1-score (Figura 6.6) e um dos treinamentos mais estáveis (Figura 6.7) com baixa variação do intervalo de confiança.

Avaliação Quantitativa

A Tabela 6.7 e a Tabela 6.8 apresentam os resultados das métricas *Recall* e F1-score, respectivamente, nas imagens de teste do conjunto de dados de Eventos Naturais no Mar. O modelo Xception atingiu os maiores valores das duas métricas quase para todas as classes, demonstrando os benefícios de empregar múltiplos blocos de Inception. Por outro lado, o uso de conexões residuais no modelo Inception ResNet não trouxe melhorias significativas comparado com o modelo base, Inception. Finalmente, podemos perceber que o uso de mecanismos de atenção no modelo DAS trouxe ganhos significativos (até 0.08 e 0.04 em termos de *Recall* e F1-score, respectivamente, para a classe Frente oceânica) comparado com o modelo base sem o uso de atenção (VGG16).

Tabela 6.7: Resultados da métrica *Recall* dos modelos avaliados no conjunto de teste das imagens de Eventos naturais no Mar.

Classe	Incep	Xcep	Incep-R	MNet	VGG16	DAS
Ondas puras do oceano	0.98	0.99	0.98	0.95	0.98	0.98
Rajada de vento	0.96	0.98	0.96	0.90	0.96	0.97
Células micro-convectivas	0.98	0.99	0.98	0.98	0.96	0.98
Células de chuva	0.97	0.89	0.94	0.99	0.95	0.97
Manchas biológicas	0.92	0.98	0.98	0.99	0.93	0.98
Gelo marinho	0.95	0.96	0.89	0.98	0.94	0.97
Icebergs	0.96	0.91	0.94	0.97	0.95	0.96
Área de vento fraco	0.94	0.97	0.93	0.96	0.92	0.94
Frente atmosférica	0.93	0.94	0.92	0.97	0.96	0.92
Frente oceânica	0.88	0.92	0.84	0.86	0.75	0.83

Tabela 6.8: Resultados da métrica F1-score dos modelos avaliados no conjunto de teste das imagens de Eventos naturais no Mar.

Classe	Incep	Xcep	Incep-R	MNet	VGG16	DAS
Ondas puras do oceano	0.94	0.96	0.92	0.94	0.89	0.91
Rajada de vento	0.97	0.97	0.97	0.90	0.94	0.93
Células micro-convectivas	0.97	0.98	0.96	0.94	0.95	0.97
Células de chuva	0.97	0.95	0.96	0.98	0.91	0.94
Manchas biológicas	0.93	0.94	0.93	0.96	0.93	0.92
Gelo marinho	0.98	0.99	0.91	0.96	0.92	0.96
Icebergs	0.93	0.98	0.92	0.99	0.94	0.95
Área de vento fraco	0.96	0.97	0.91	0.95	0.89	0.93
Frente atmosférica	0.95	0.96	0.94	0.95	0.94	0.93
Frente oceânica	0.89	0.95	0.87	0.88	0.73	0.77

Avaliação Qualitativa

Como foi apresentado no Capítulo 3, o GradCAM é usado para visualizar as áreas de maior interesse de acordo com o modelo para gerar a predição.

A Figura 6.8 apresenta os resultados da inferência no nível do *patch* para os dados SAR disponíveis usando os modelos avaliados. De esquerda a direita se apresentam a imagem SAR, e os mapas de ativação para o resultado da inferência em cada modelo e de cima para baixo, os resultados da inferência para cada classe serão apresentados usando os vários modelos estudados.

O estudo dos mapas de ativação permite ter uma noção importante do que a rede neural convolucional está aprendendo. No conjunto de dados de Eventos Naturais no Mar temos dois tipos de classes: as classes definidas pela imagem toda (e.g. classe Ondas puras do oceano), e as classes definidas por uma região da imagem (e.g. classe Icebergs).

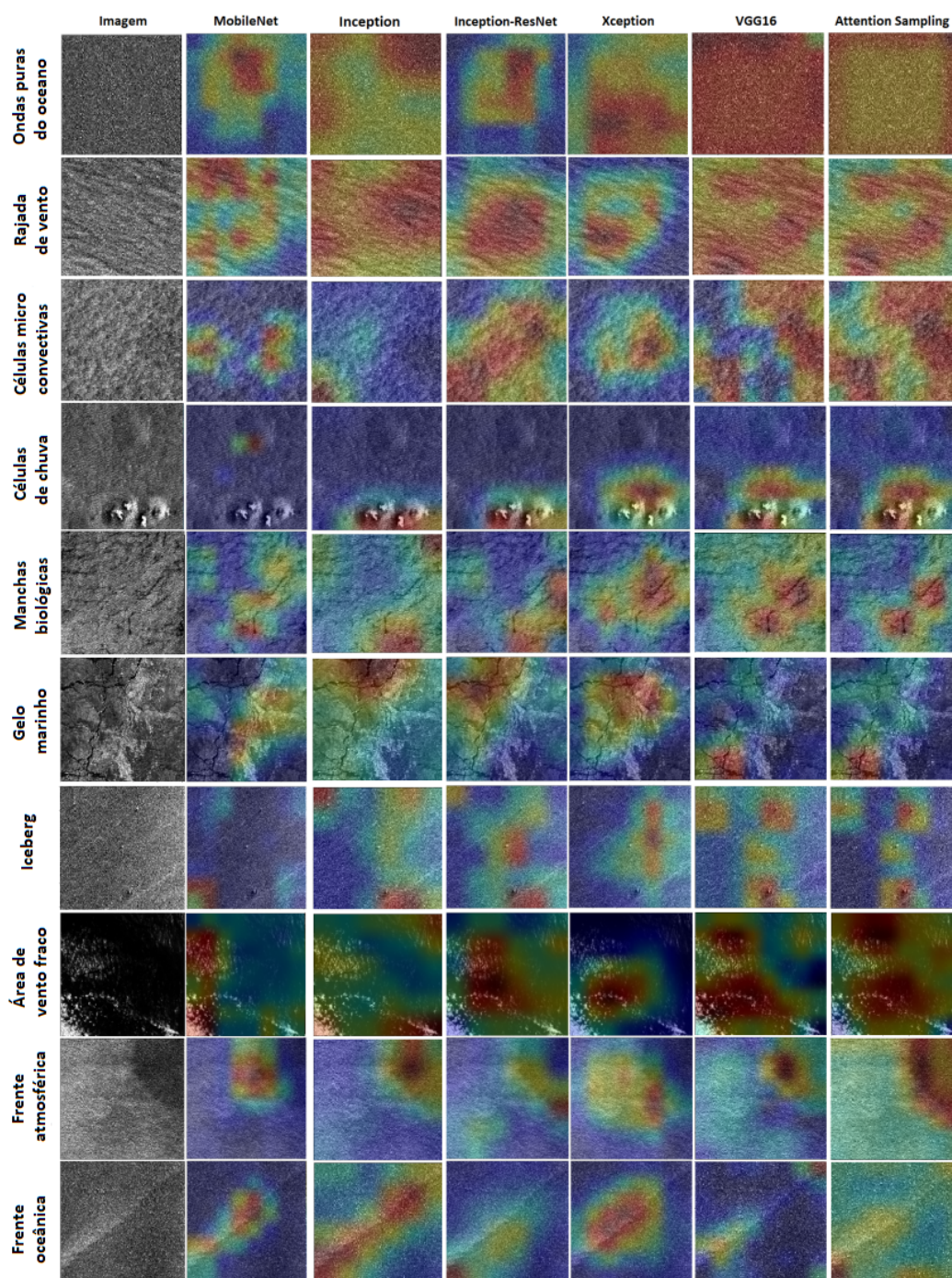


Figura 6.8: GradCAM dos modelos avaliados no conjunto de dados de Eventos Naturais no Mar.

Para as classes Células de chuva, Manchas biológicas, *Iceberg*, Frente oceânica, e Frente atmosférica, pode-se observar que de acordo o GradCAM (Figura 6.8) a classe de interesse fica presente em pequenas regiões da imagem. Este efeito pode ser mais explícito na classe *Iceberg*, onde pequenos pedaços de gelo estão presentes na superfície do mar. No caso das classes Frente atmosférica e Frente oceânica, esse efeito estará relacionado ao fato de a

classe só poder ser representada por uma linha descrevendo 2 correntes de ar ascendentes e descendentes ou limitante que separa 2 gradientes de intensidade, respectivamente.

Em comparação com as classes Onda pura do oceano, Células microconvectivas e Área de vento fraco, pode-se observar como essas cenas conterão uma maior quantidade de informações devido à natureza do fenômeno. Um exemplo disso pode ser a classe Onda pura do oceano, onde a partir de uma incidência de uma corrente de vento ascendente ou descendente, podem ser geradas variações periódicas na superfície, mostrando uma mudança na rugosidade sobre uma área extensa.

Com base nisso, os modelos Xception e Inception ResNet apresentaram bom desempenho mostrando mapas de ativação nas áreas mais relevantes da cena. Para o caso da Inception ResNet, na classe Frente atmosférica pôde-se constatar que o mapa de ativação representou a área com maior retroespalhamento, tendo uma diminuição de importância no limite entre dois gradientes de retroespalhamento. Para o Xception ficou evidente que a importância estava na área escurecida onde os 2 gradientes são encontrados, indicando, portanto, que uma pequena parte desse limite descreve o evento.

Em relação ao desempenho da arquitetura VGG16 vs. Deep Attention Sampling, foi evidenciado pela análise GradCAM que para as classes Frente oceânica e Frente atmosférica, o Deep Attention Sampling teve um foco mais centralizado em algumas regiões da imagem, mostrando uma clara diferença na aprendizagem de atributos por classe, onde Deep Attention Sampling mostrou um mapa de ativação mais próximo da definição do evento.

Em termos gerais, foi evidenciado para este conjunto de dados que as arquiteturas baseadas no Inception apresentaram um bom desempenho na classificação das 10 classes. Um dos fatores mais relevantes está relacionado na profundidade que essas arquiteturas conseguem ter. Isso pôde ser evidenciado quantitativa e qualitativamente, onde essas arquiteturas alcançaram uma pontuação mais elevada das métricas *Recall* e F1-score, possuindo mapas de ativação de acordo com o que era esperado para cada classe.

6.4.2

Derramamento de Óleo no Mar

Avaliação do Treinamento

A Figura 6.9 apresenta a evolução do F1-score durante o treinamento dos modelos avaliados. As curvas representa a média do F1-score para 40 rodadas

por modelo. Os modelos Inception, Xception e Deep Attention Sampling obtiveram os maiores valores de F1-score: 0.94, 0.95 e 0.96, respectivamente. Seguido, as arquiteturas VGG16, MobileNet e Inception ResNet obtiveram valores de F1-score de 0.92, 0.89 e 0.86.

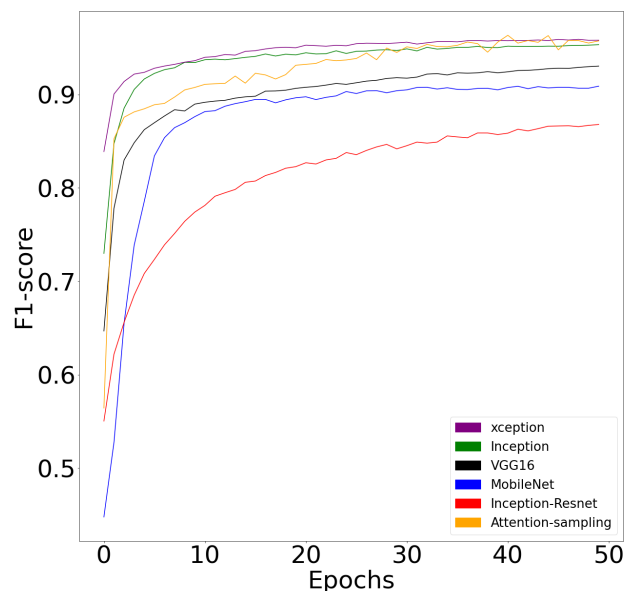


Figura 6.9: Curvas de treinamento no conjunto de dados de derramamento de óleo: F1-score vs. Épocas. O gráfico apresenta a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.

Na análise dos valores de F1-score obtido por cada modelo foi calculado o intervalo de confiança das 40 rodadas no treinamento dos modelos, a Figura 6.10 apresenta o intervalo de confiança para cada modelo. O modelo mais estável em termos de intervalo de confiança é o Xception tendo uma variação de ± 0.02 . Comparado a outras arquiteturas, Inception obteve um valor de ± 0.03 , Inception ResNet de ± 0.04 , VGG16 de ± 0.05 , MobileNet de ± 0.04 e Deep Attention Sampling de ± 0.03 .

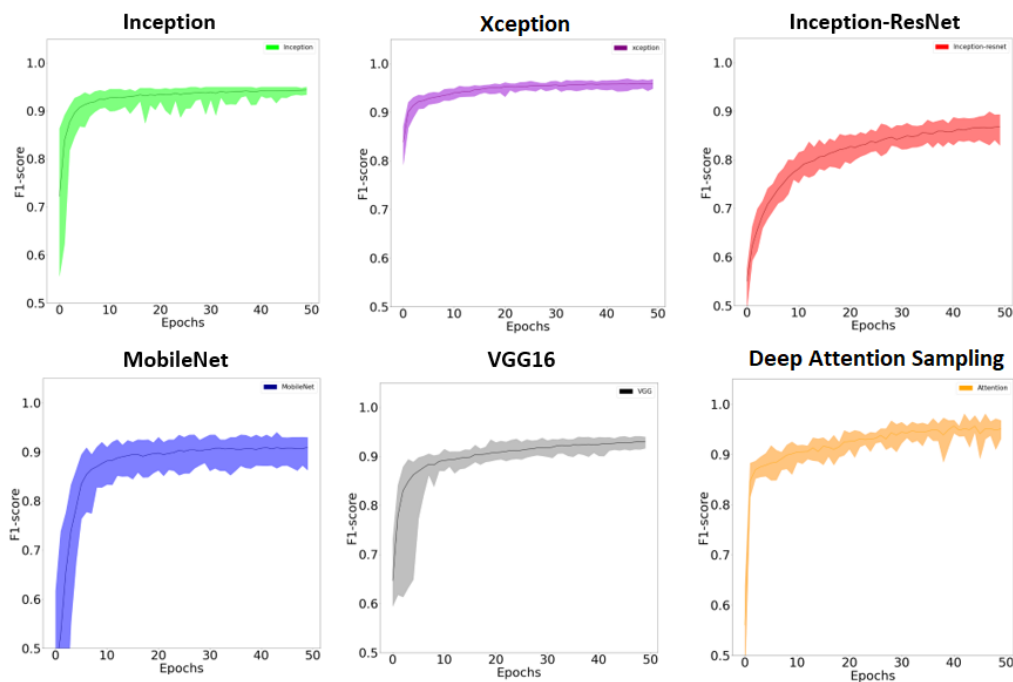


Figura 6.10: Intervalo de confiança das curvas de treinamento no conjunto de dados do Derramamento de Óleo no Mar: F1-score vs. Número de Épocas. O gráfico apresenta o intervalo de confiança para a média da métrica F1-score das 40 rodadas para cada uma das arquiteturas avaliadas.

Avaliação Quantitativa

Para o estudo de derramamento de óleo no mar, o *Recall* foi avaliado quanto à presença de 2 classes óleo e não óleo. O estudo da métrica teve como objetivo estudar a quantidade de falsos negativos (FN) presentes durante a inferência. A presença do FN (óleo classificado como não óleo) definirá o desempenho do modelo, procurando a classificação correta da classe óleo. A Tabela 6.10 apresenta a média da métrica *Recall* para cada arquitetura estudada, onde os maiores valores de *Recall* foram obtidos pelas arquiteturas Deep Attention Sampling e Xception. Nesse cenário, pode-se observar como a Deep Attention sampling torna-se melhor que o VGG16, mostrando que os módulos de atenção neste cenário têm um efeito positivo. Para este conjunto de dados, notou-se principalmente que os *patches* foram extraídos baseados no número de pixels da classe presente na cena. Considerando isso, ter um critério baseado na atenção ajudou a obter uma pontuação mais elevada nas métricas de interesse, isto baseado principalmente na obtenção de representações com uma presença mais significativa da classe.

Como é apresentado na Tabela 6.10, a classificação da classe derramamento de óleo apresentou um grande desafio na grande maioria dos modelos, principalmente porque a cena continha outros eventos presentes, os quais tinham valores próximos ao derramamento de óleo. Um exemplo claro disso foi a possível presença de *low wind areas*.

Tabela 6.9: Resultados da métrica F1 score dos modelos avaliados no conjunto de teste das imagens de Derramamento de Óleo no Mar.

Classe	Incep	Xcep	Incep-R	MNet	VGG16	DAS
Não óleo	0.87	0.90	0.70	0.87	0.90	0.92
Óleo	0.76	0.80	0.67	0.73	0.63	0.82

Tabela 6.10: Resultados da métrica Recall dos modelos avaliados no conjunto de teste das imagens de Derramamento de Óleo no Mar.

Classe	Incep	Xcep	Incep-R	MNet	VGG16	DAS
Não óleo	0.92	0.94	0.95	0.95	0.96	0.93
Óleo	0.79	0.80	0.68	0.74	0.64	0.87

As arquiteturas previamente estudadas foram validadas em um conjunto de dados privado com imagens do satélite Cosmos SkyMed. O objetivo era utilizar um conjunto de dados sem pré-processamento e com uma resolução espacial diferente, tendo foco nas arquiteturas Xception e Deep Attention Sampling. A Tabela 6.4.2 mostra o *Recall* obtido na inferência, demonstrando como a abordagem estudada pode ser implementada em outros tipos de problemas.

Tabela 6.11: Resultados da métrica Recall no conjunto de teste com dados confidenciais

Classe	Xcep	DAS
0	0.92	0.96
1	0.68	0.73

Avaliação Qualitativa

A Figura 6.11 apresenta os resultados da inferência de uma imagem SAR obtida no 19/15/2018 usando os modelos avaliados.

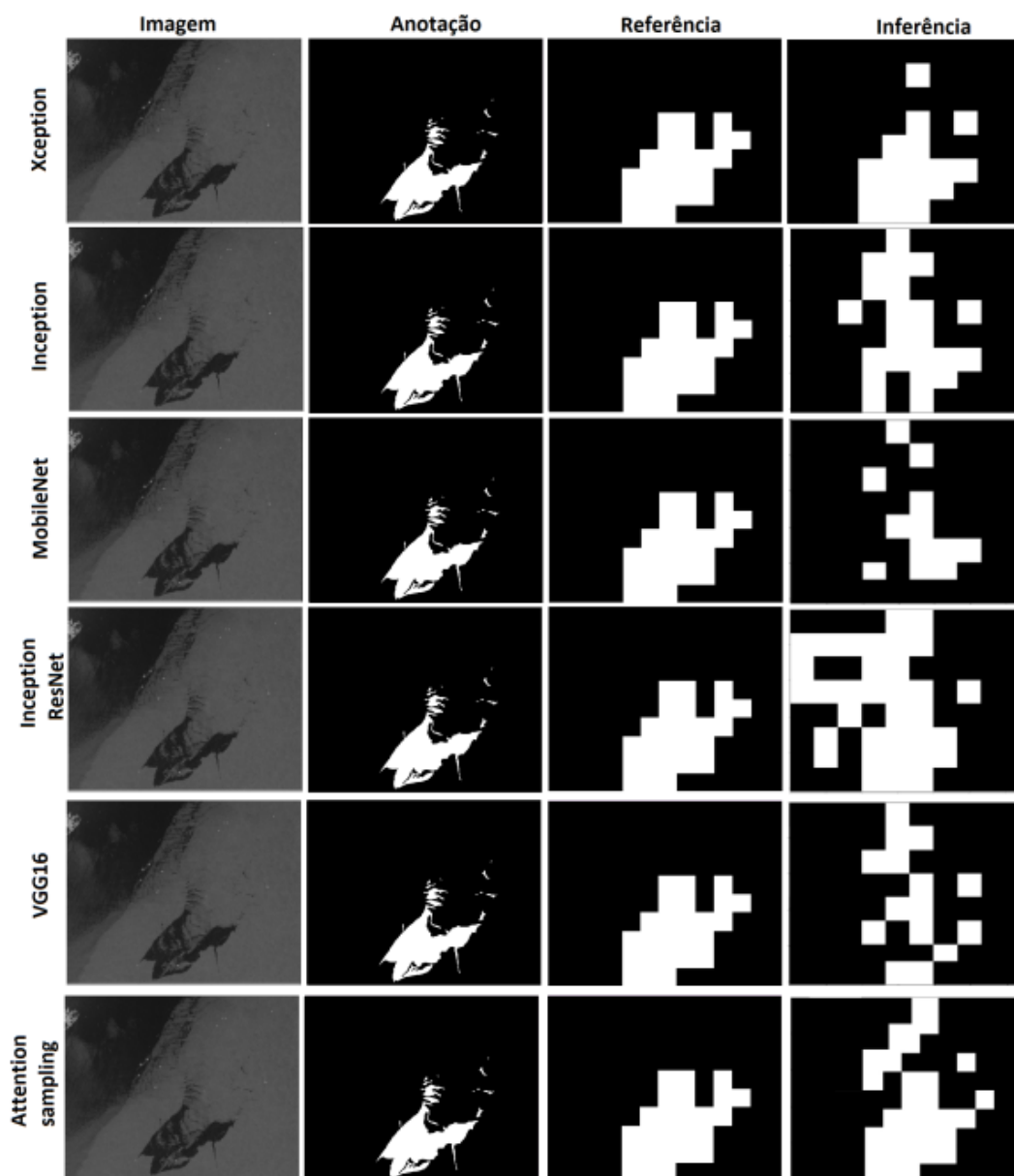


Figura 6.11: Resultado da inferência de uma imagem SAR (19/12/2018) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de *patches*, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).

De esquerda a direita a Figura 6.11 apresenta a Anotação e Referência (etiquetas a nível de pixel e recorte, respectivamente), e a Inferência (o mapa de classificação) com as classes óleo (branco) e não óleo (preto). Xception e Deep Attention Sampling obtiveram resultados mais semelhantes a referência com uma menor quantidade de falsos negativos. Na imagem SAR podemos perceber a presença de outra classe com um retroespalhamento próximo ao de óleo (áreas

escuras), a qual poderia ser uma área de baixo vento. A Inception ResNet teve problemas para distinguir essas classes, gerando uma maior quantidade de falsos positivos.

Na Figura 6.12 são apresentados os resultados da inferência de outra imagem SAR obtida no 19/12/2018 usando os modelos avaliados. Cada coluna segue o mesmo padrão da Figura 6.11. Nesta imagem podemos observar a presença de áreas levemente escuras no mar que poderiam representar uma frente oceânica. O Deep Attention Sampling gerou um mapa de classificação mais próximo da referência tendo um resultado melhor que VGG16, demonstrando as vantagens de usar mecanismos de atenção para o treinamento da mesma rede base - VGG16.

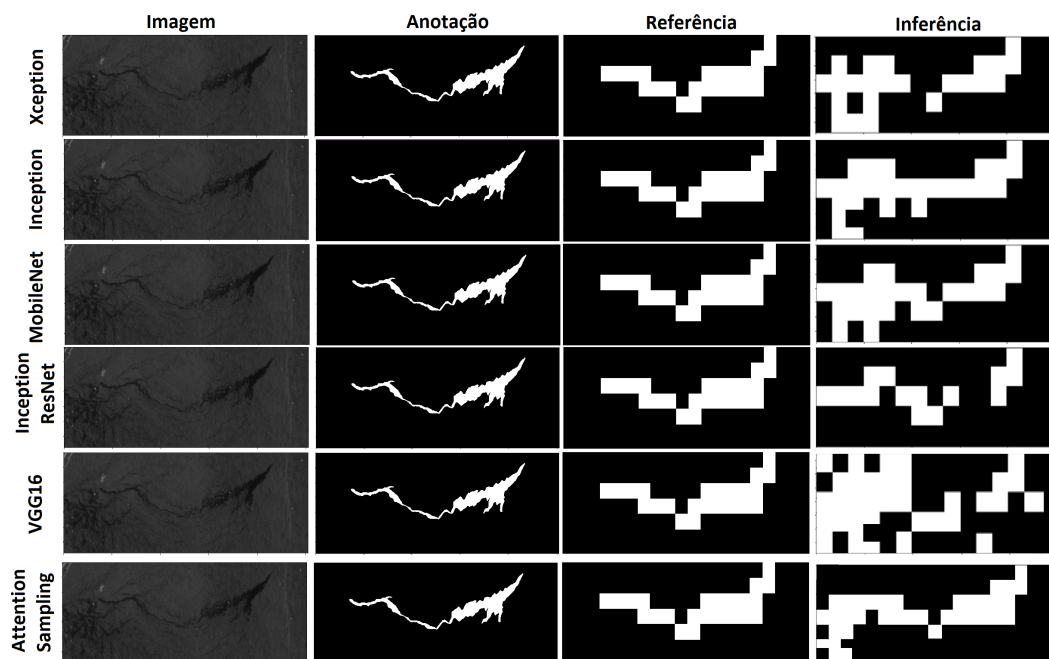
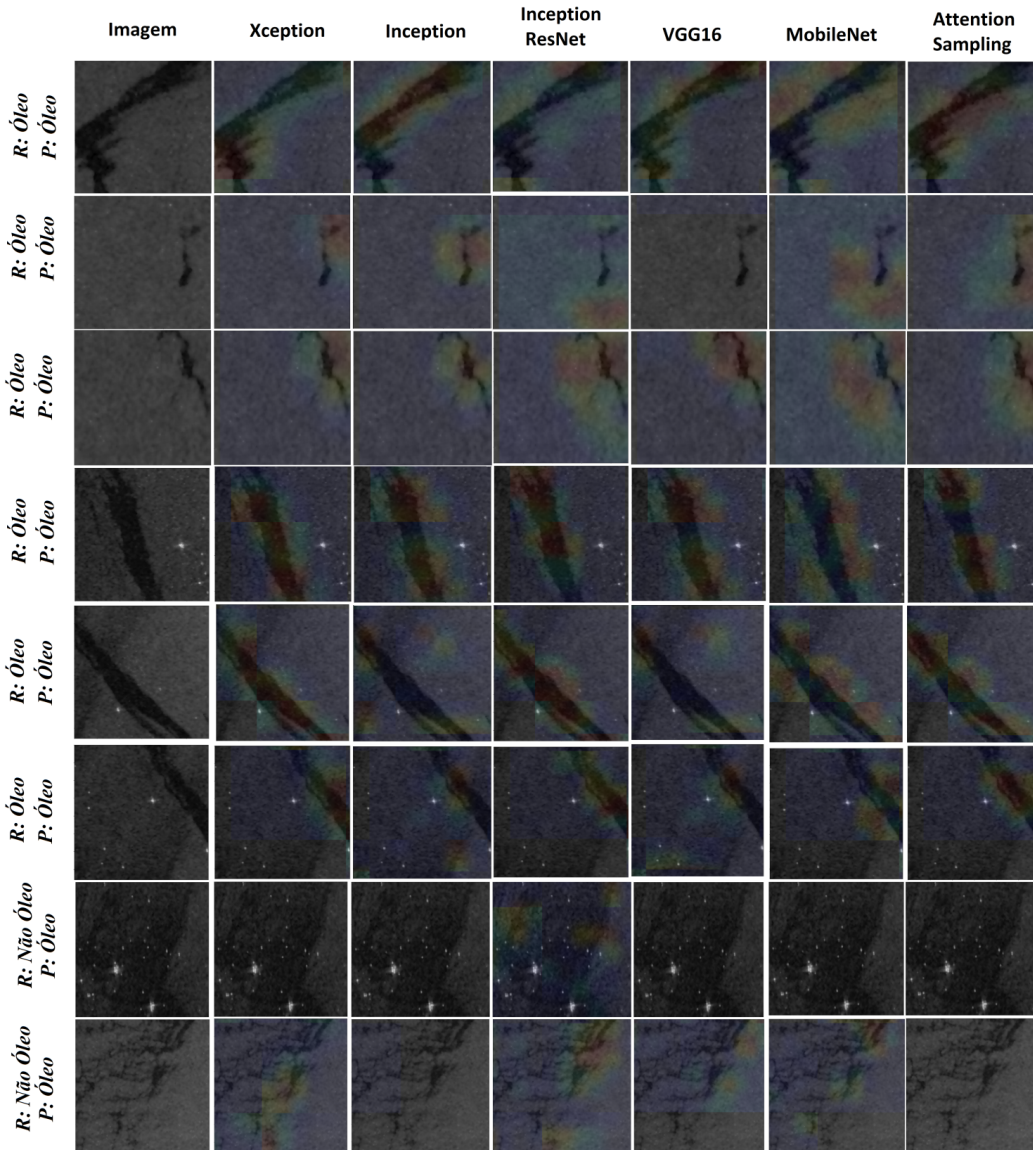


Figura 6.12: Resultado da inferência de uma imagem SAR (15/10/2019) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de *patches*, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).

Em relação a outras arquiteturas como a Inception ResNet, foi observada uma grande dificuldade na classificação de áreas da classe não óleo com retroespalhamento próximo à classe de derramamento de óleo.

Baseado no GradCAM, pode-se observar que o modelo obtido de Deep Attention sampling teve mapas de ativação mais próximos do derramamento de óleo, descrevendo melhor a classe, como pode ser visto na Figura 6.13.

Em comparação com o modelo usando VGG16, pode-se afirmar que a Deep Attention sampling consegue relacionar de forma mais precisa as áreas com representações mais finas.



PUC-Rio - Certificação Digital Nº 1912831/CA

Figura 6.13: *patches* das imagens SAR de teste no conjunto de dados de Derramamento de Óleo no Mar. O GradCAM apresenta o mapa de ativação para inferência, onde a matriz de ativação é sobreposta na imagem do estudo.

Neste cenário as áreas com a menor representação no nível de pixel da classe óleo, Deep Attention Sampling consegue definir um mapa de ativação de forma correta centralizada no derramamento de óleo.

No Apêndice A se apresentam os resultados da inferência de outras imagens SAR no conjunto de teste usando os modelos avaliados. Na Figura A.4 um derramamento de óleo é observado, tendo a presença de valores de

retroespalhamento próximos ao derramamento de óleo na classe não óleo. Visualmente, a cena representa um cenário altamente complexo onde as áreas escuras não podem ser distinguidas respeito da classe de derramamento de óleo. Desta forma, observou-se que para os modelos de Xception e Deep Attention sampling, algumas dessas regiões sem óleo foram classificadas como óleo.

A Figura A.3 mostra outra cena onde é possível observar por inspeção visual a presença de derramamento de óleo. Nesse cenário é possível observar como a Deep Attention Sampling consegue ter uma inferência mais próxima da referência, tendo um cenário com pouca variabilidade no retroespalhamento.

A partir da Figura A.3 pode-se observar que, o modelo obtido para Deep Attention sampling apresentou um bom desempenho frente às variações de retroespalhamento em torno da área de interesse (derramamento de óleo). Isso foi observado ao estudar a cena, onde a classe não óleo descreveu outro possível evento no mar. Neste cenário é observando como o modelo tem uma grande suscetibilidade a variações quanto à incidência de vento e variação na rugosidade da superfície.

7

CONCLUSÕES

Se apresentou uma avaliação das arquiteturas de aprendizado profundo baseadas em redes neurais convolucionais para classificação de imagens aplicado ao monitoramento de áreas marinhas para a detecção de feições de interesse. Os modelos avaliados estão baseados em camadas convolutivas (VGG16, MobileNet), módulos Inception (Inception, Xception), conexões residuais (Inception-ResNet) e mecanismos de atenção (Deep Attention Sampling). Foram usados dois conjuntos de dados com imagens SAR representando dois tipos de eventos no mar: Eventos Naturais no Mar (10 classes, balanceado), e Derramamento de Óleo no Mar (2 classes, desbalanceado). Para a avaliação dos modelos foi feita uma análise do treinamento de cada modelo (curva de aprendizado e intervalo de confiança), uma análise quantitativa (métricas *Recall* e F1-score), e análise qualitativa (GradCAM).

Para o conjunto de dados de Eventos Naturais no Mar, o modelo Xception apresentou os melhores resultados em termos de F1-score e *Recall* com valores de 0.96 e 0.95 (Média ponderada 10 classes), respectivamente. Também foi um dos modelos mais estáveis em termos de intervalo de confiança com uma variação de ± 0.04 . Em relação à análise quantitativa, o GradCAM apresentou mapas de ativação próximos às áreas mais importantes na descrição da classe, evidenciando o aprendizado correto.

Para o conjunto de dados derramamento de óleo, o modelo Deep Attention sampling apresentou os melhores resultados em termos de F1-score e *Recall* com valores de 0.82 e 0.87, respectivamente. Em relação aos resultados, observou-se através do mapa de inferência que o modelo Deep Attention sampling teve um bom desempenho, havendo a correta classificação das áreas de presença de derramamentos de óleo.

A partir da avaliação dos modelos VGG16 e Deep Attention sampling, o efeito positivo da implementação dos módulos de atenção pôde ser evidenciado. Com base na Deep Attention sampling, onde é implementada uma rede totalmente conectada (VGG16), isto para a extração de atributos, um fator de melhoria de um fator maior ao 18% foi evidenciado para a classe de derramamento de óleo no *Recall*. Em comparação, foi observado como a extração de recortes com base na atenção espacial pode auxiliar no processo

de treinamento, isto levando em consideração que o critério de criação de recortes foi executado através da quantidade mínima de pixels da classe na cena. Em relação às arquiteturas baseadas no Inception, uma clara diferença ficou evidente devido ao foco de cada uma delas. Para este caso, foi evidenciado como atingir uma maior profundidade na implementação do Xception gerou um melhor desempenho com base na métrica *Recall*. Isso foi observado na implementação do Inception ResNet, onde foi obtido um desempenho inferior no processo de classificação em comparação com o Xception tendo 0.68 e 0.80 no *Recall* respectivamente.

Da execução deste trabalho pode-se observar a grande importância do estudo dos dados SAR, destacando os focos que o estudo da CNN pode proporcionar em torno da área de classificação de eventos no mar. Com base nisso, para a execução de trabalhos futuros, uma abordagem mais ampla pode ser considerada quanto à análise de dados SAR. Em relação ao estudo dos modelos baseados nas CNN's, observou-se que considerar o Deep Attention sampling ajudou na classificação de eventos no mar para cenas com pouca representação do evento. Com base nisso, a possibilidade de incluir uma rede mais profunda na implementação da Deep Attention sampling pode ajudar a obter um modelo de classificação com um desempenho superior.

Referências bibliográficas

- [1] MAGRIS, R. A.; GIARRIZZO, T.. **Mysterious oil spill in the Atlantic Ocean threatens marine biodiversity and local people in Brazil.** *Marine Pollution Bulletin*, 153:110961, 2020.
- [2] IVANOVA, M.. **Oil spill emergency preparedness in the Russian Arctic: a study of the Murmansk region.** *Polar Research*, 30(1):7285, 2011.
- [3] KLAPP, J.; CROS, A.; FUENTES, O. V.; STERN, C. ; MEZA, M. A. R.. **Experimental and Theoretical Advances in Fluid Dynamics.** Springer, 2012.
- [4] JEON, I.; HAM, S.; CHEON, J.; KLIMKOWSKA, A. M.; KIM, H.; CHOI, K. ; LEE, I.. **a Real-Time Drone Mapping Platform for Marine Surveillance.** *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4213:385–391, June 2019.
- [5] WÖLFL, A.-C.; JENCKS, J.; JOHNSTON, G.; VARNER, J. ; DEVEY, C.. **Identifying deep-sea target areas for a pilot Atlantic seabed mapping project using GIS techniques.** *The Journal of Ocean Technology*, 12(4):28–42, 2017.
- [6] WANG, C.; MOUCHE, A.; TANDEO, P.; STOPA, J.; LONGÉPÉ, N.; ERHARD, G.; FOSTER, R.; VANDEMARK, D. ; CHAPRON, B.. **Labeled SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode (TenGeoP-SARwv).** *Geoscience Data Journal*, 2018.
- [7] TONG, S.; LIU, X.; CHEN, Q.; ZHANG, Z. ; XIE, G.. **Multi-Feature Based Ocean Oil Spill Detection for Polarimetric SAR Data Using Random Forest and the Self-Similarity Parameter.** *Remote Sensing*, 11(4), 2019.
- [8] GARCIA-PINEDA, O.; MACDONALD, I. R.; LI, X.; JACKSON, C. R. ; PICHEL, W. G.. **Oil Spill Mapping and Measurement in the Gulf of Mexico With Textural Classifier Neural Network Algorithm**

- (TCNNA). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(6):2517–2525, 2013.
- [9] TAN, W.; LI, J.; XU, L. ; CHAPMAN, M. A.. **Semiautomated segmentation of sentinel-1 sar imagery for mapping sea ice in labrador coast.** *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1419–1432, 2018.
- [10] CHEN, G.; LI, Y.; SUN, G. ; ZHANG, Y.. **Application of Deep Networks to Oil Spill Detection Using Polarimetric Synthetic Aperture Radar Images.** *Applied Sciences*, 7(10), 2017.
- [11] JOSEPH, M.; JAYASRI, P. V.; DUTTA, S.; KUMARI, E. V. S. S. ; PRASAD, A. V. V.. **Oil spill detection from RISAT-1 imagery using texture analysis.** In: 2016 Asia-Pacific Microwave Conference (APMC), p. 1–4, 2016.
- [12] XU, J.; PAN, X.; JIA, B.; WU, X.; LIU, P. ; LI, B.. **Oil Spill Detection Using LBP Feature and K-Means Clustering in Shipborne Radar Image.** *Journal of Marine Science and Engineering*, 9(1), 2021.
- [13] AL-RUZOUQ, R.; GIBRIL, M. B. A.; SHANABLEH, A.; KAIS, A.; HAMED, O.; AL-MANSOORI, S. ; KHALIL, M. A.. **Sensors, Features, and Machine Learning for Oil Spill Detection and Monitoring: A Review.** *Remote Sensing*, 12(20), 2020.
- [14] SU, N.; CHEN, X.; GUAN, J. ; LI, Y.. **Deep CNN-Based Radar Detection for Real Maritime Target Under Different Sea States and Polarizations.** In: *International Conference on Cognitive Systems and Signal Processing*, p. 321–331. Springer, 2018.
- [15] YAOHUA, X.; XUDONG, M.. **A SAR Oil Spill Image Recognition Method Based on DenseNet Convolutional Neural Network.** In: 2019 International Conference on Robots Intelligent System(ICRIS), p. 78–81, 2019.
- [16] BOULZE, H.; KOROSOV, A. ; BRAJARD, J.. **Classification of Sea Ice Types in Sentinel-1 SAR Data Using Convolutional Neural Networks.** *Remote Sensing*, 13(13), 2020.
- [17] LI, Y.; YANG, X.; YE, Y.; CUI, L.; JIA, B.; JIANG, Z. ; WANG, S.. **Detection of Oil Spill Through Fully Convolutional Network.** In: *International Conference on Geo-Spatial Knowledge and Intelligence*, 2017.

- [18] WANG, C.; TANDEO, P.; MOUCHE, A.; STOPA, J. E.; GRESSANI, V.; LONGEPE, N.; VANDEMARK, D.; FOSTER, R. C. ; CHAPRON, B.. **Classification of the global Sentinel-1 SAR vignettes for ocean surface process studies**. *Remote Sensing of Environment*, 234:111457, 2019.
- [19] KATHAROPOULOS, A.; FLEURET, F.. **Processing Megapixel Images with Deep Attention-Sampling Models**. In: *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.
- [20] LILLESAND, T.; KIEFER, R. W. ; CHIPMAN, J.. **Remote Sensing and Image Interpretation**. John Wiley & Sons, 2015.
- [21] VAN ZYL, J. J.. **Synthetic Aperture Radar Polarimetry**, volumen 2. John Wiley & Sons, 2011.
- [22] HOWELL, S. E.; KOMAROV, A. S.; DABBOOR, M.; MONTPETIT, B.; BRADY, M.; SCHARIEN, R. K.; MAHMUD, M. S.; NANDAN, V.; GELDSSETZER, T. ; YACKEL, J. J.. **Comparing l-and c-band synthetic aperture radar estimates of sea ice motion over different ice regimes**. *Remote Sensing of Environment*, 204:380–391, 2018.
- [23] LECUN, Y.; BOTTOU, L.; BENGIO, Y. ; HAFFNER, P.. **Gradient-based learning applied to document recognition**. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [24] KRIZHEVSKY, A.; SUTSKEVER, I. ; HINTON, G. E.. **Imagenet classification with deep convolutional neural networks**. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [25] SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D. ; BATRA, D.. **Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization**. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, p. 618–626, 2017.
- [26] SIMONYAN, K.; ZISSERMAN, A.. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. In: Bengio, Y.; LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [27] SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V. ; RABINOVICH, A.. **Going deeper with convolutions**. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 1–9, 2015.

- [28] CHOLLET, F.. **Xception: Deep learning with depthwise separable convolutions**. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 1251–1258, 2017.
- [29] SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V. ; ALEMI, A.. **Inception-v4, inception-resnet and the impact of residual connections on learning**. In: Proceedings of the AAAI Conference on Artificial Intelligence, volumen 31, 2017.
- [30] HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M. ; ADAM, H.. **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**. Computing Research Repository, abs/1704.04861, 2017.
- [31] LI, H.; MOUCHE, A.; STOPA, J. E. ; CHAPRON, B.. **Calibration of the Normalized Radar Cross Section for Sentinel-1 Wave Mode**. IEEE Transactions on Geoscience and Remote Sensing, 57(3):1514–1522, 2019.
- [32] FILIPPONI, F.. **Sentinel-1 GRD Preprocessing Workflow**. 3rd International Electronic Conference on Remote Sensing, 18(1), 2019.

Apêndices

A

Resultados de Inferência do conjunto de dados de Derramamento de Óleo no mar.

Esse apêndice mostra os resultados da inferência para as arquiteturas estudadas, usando o conjunto de dados de derramamento de óleo.

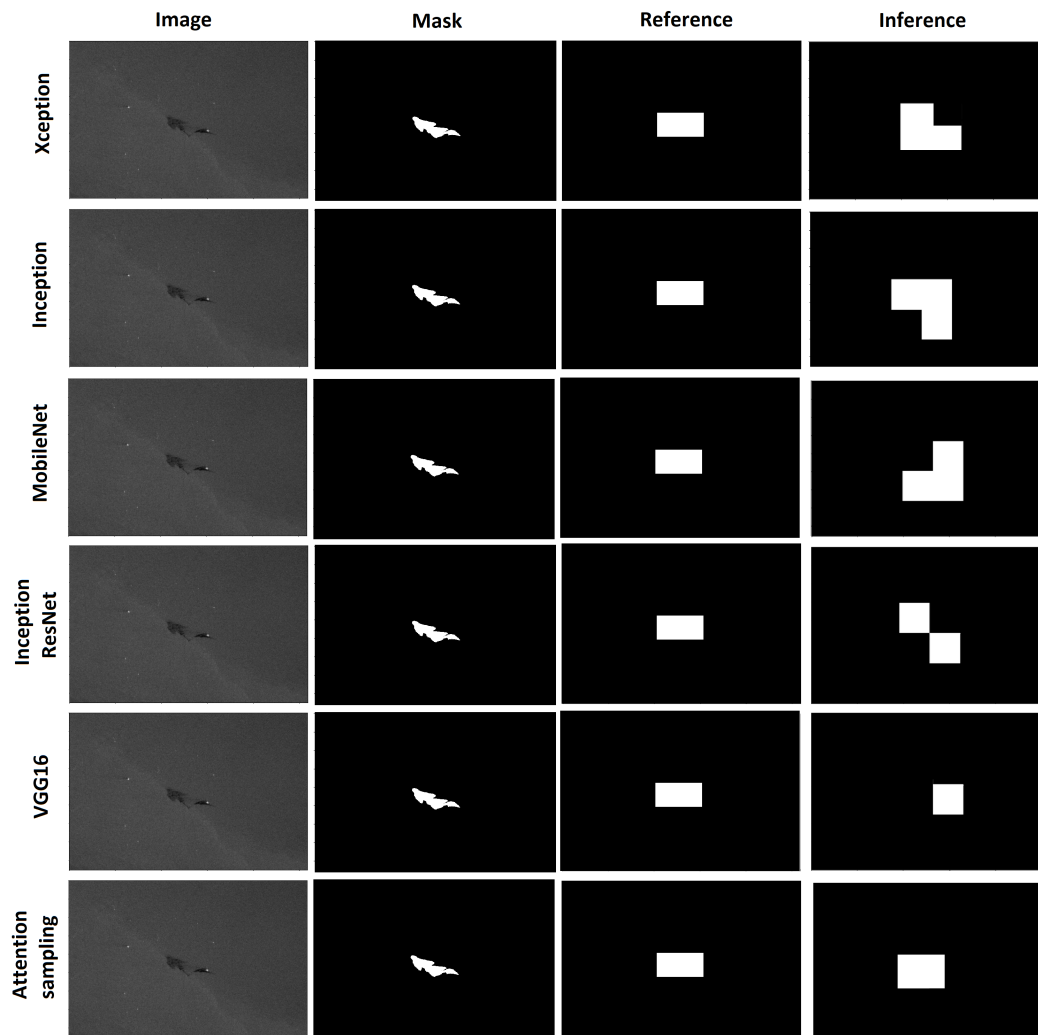


Figura A.1: Resultado da inferência de uma imagem SAR (24/02/2020) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).

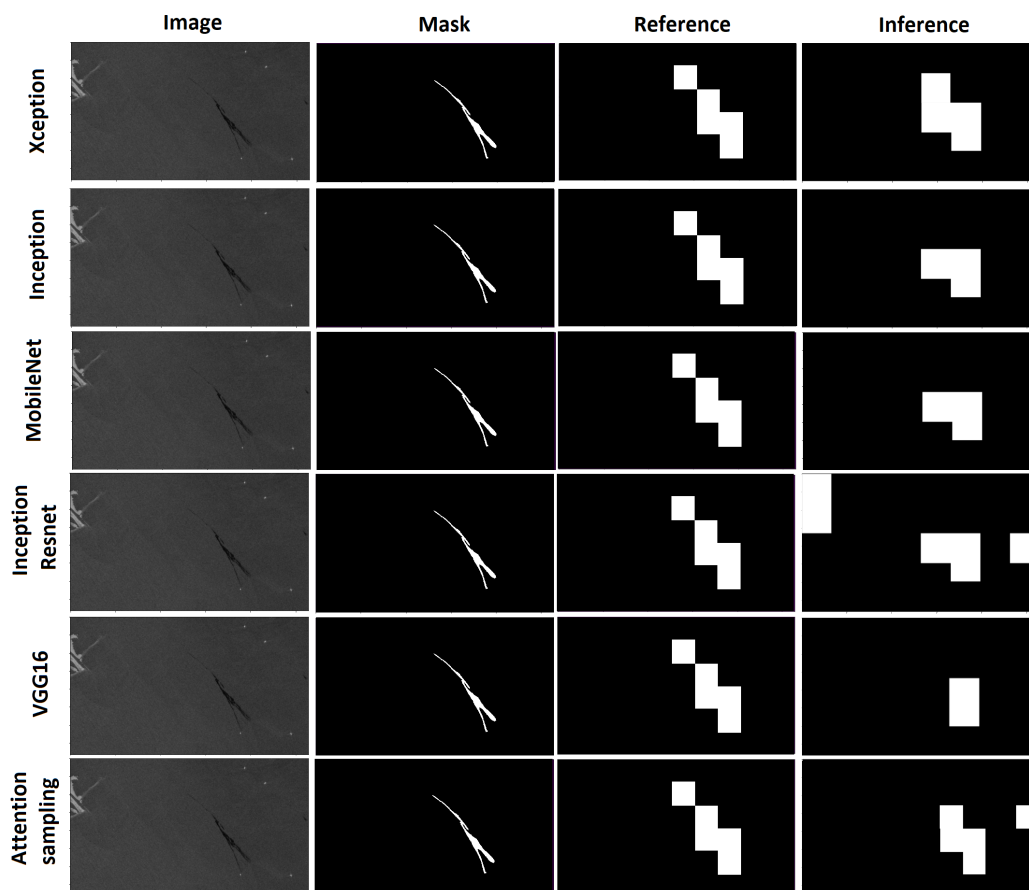


Figura A.2: Resultado da inferência de uma imagem SAR (19/03/2020) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).

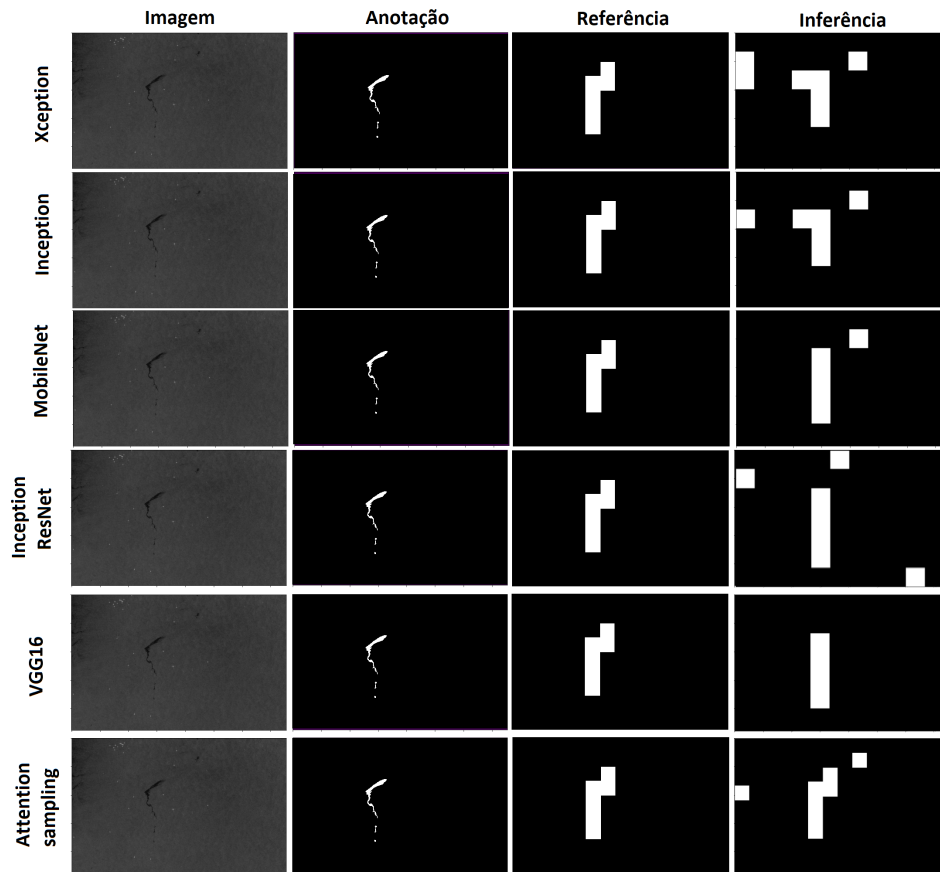


Figura A.3: Resultado da inferência de uma imagem SAR (19/03/2018) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).

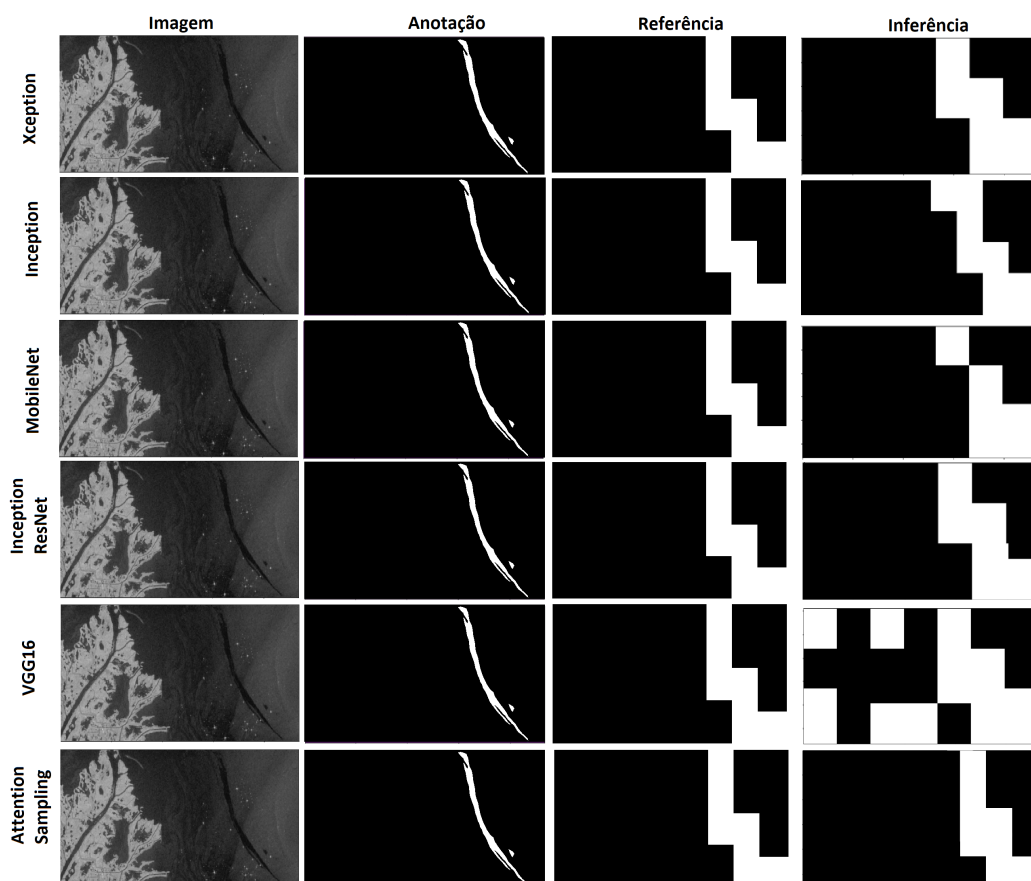


Figura A.4: Resultado da inferência de uma imagem SAR (19/12/2018) usando os modelos avaliados. A Anotação e a Referência são as etiquetas a nível de pixel e de recortes, respectivamente. Na inferência observam-se as áreas com óleo (branco) e sem óleo (preto).

B
Resultados de Inferência do conjunto de dados de eventos naturais no mar.

Esse apêndice mostra os resultados da inferência para as arquiteturas estudadas, usando o conjunto de dados de eventos naturais no mar.

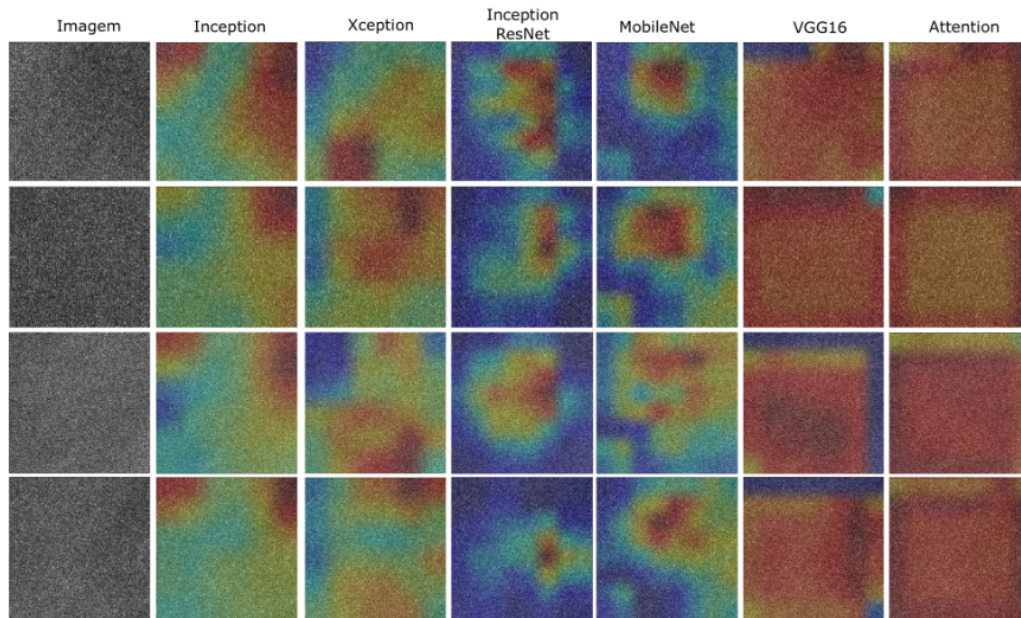


Figura B.1: GradCAM para amostras da classe Ondas puras do oceano.

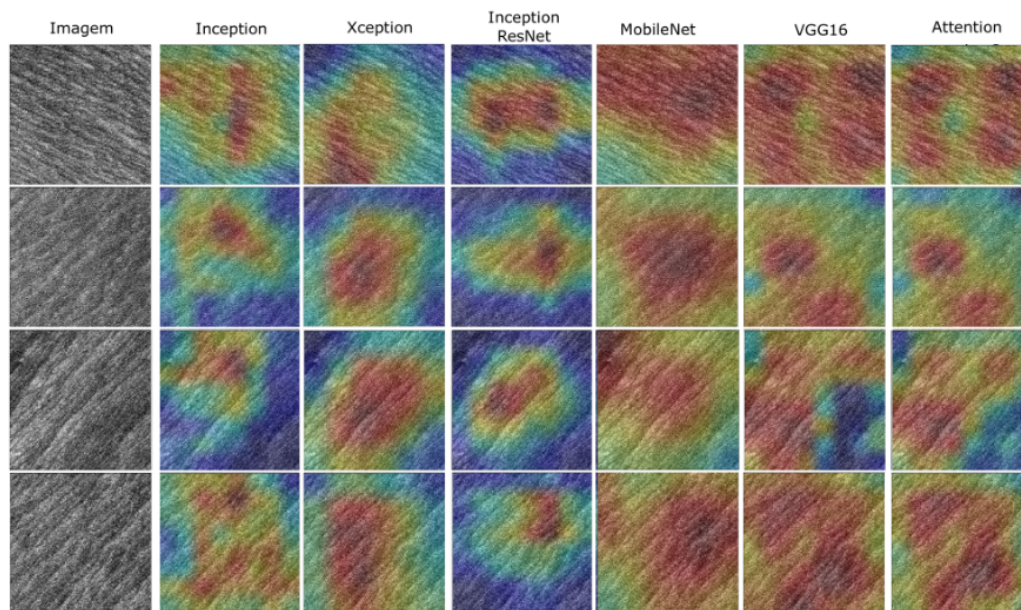


Figura B.2: GradCAM para amostras da classe Rajadas de vento.

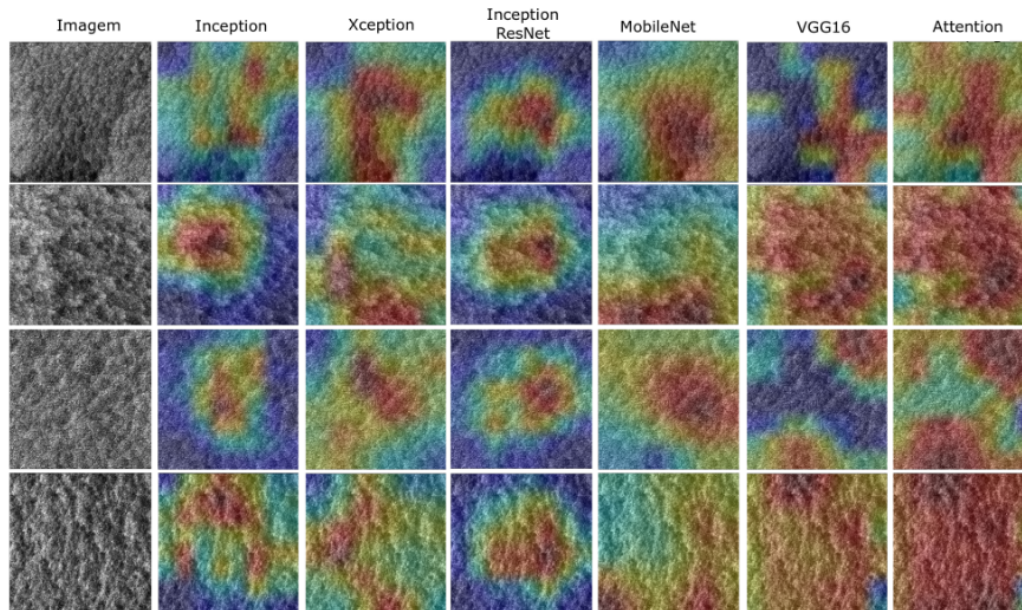


Figura B.3: GradCAM para amostras da classe Células Micro-convectivas.

PUC-Rio - Certificação Digital Nº 1912831/CA

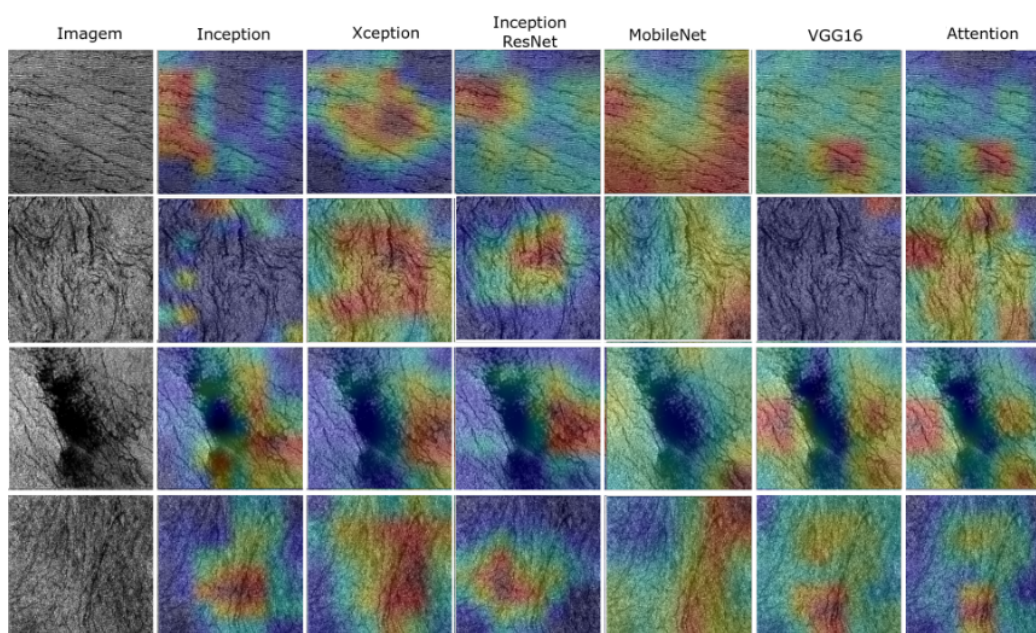


Figura B.4: GradCAM para amostras da classe Manchas biológicas.

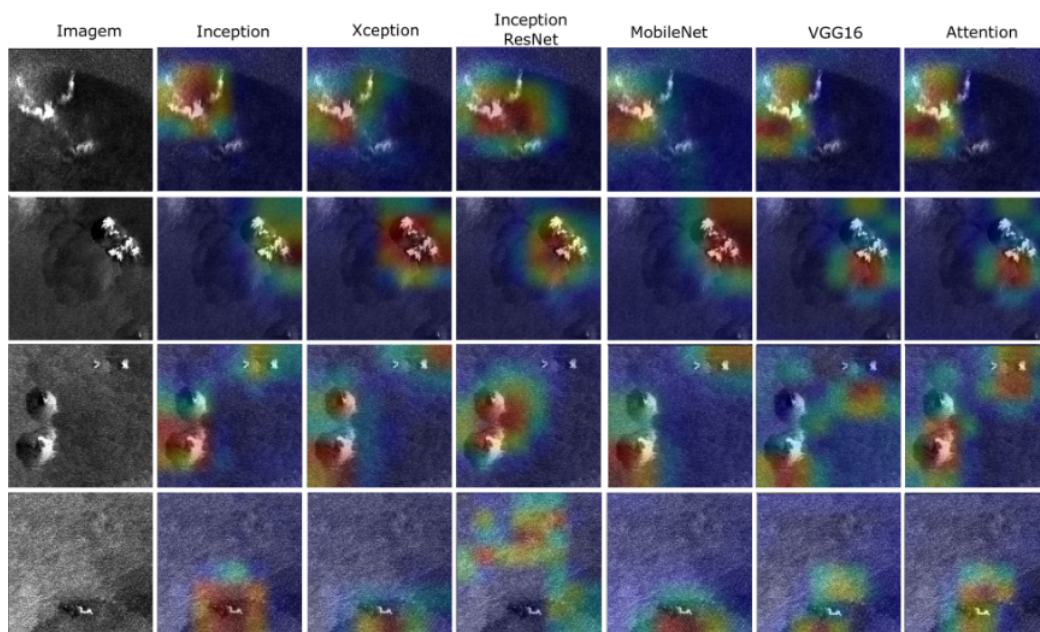


Figura B.5: GradCAM para amostras da classe Células de chuva.

PUC-Rio - Certificação Digital Nº 1912831/CA

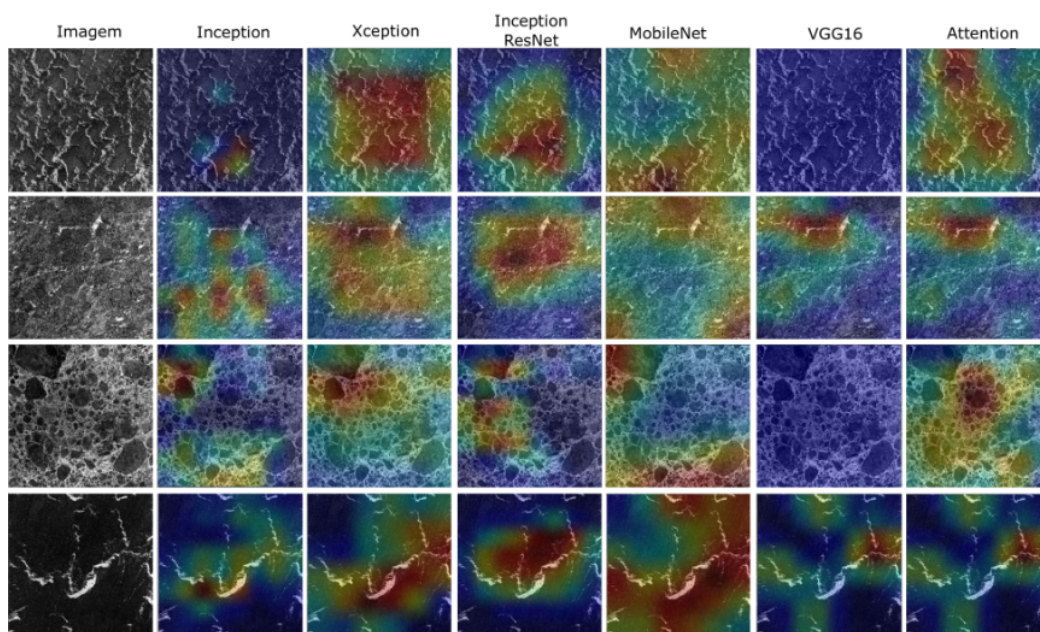


Figura B.6: GradCAM para amostras da classe Gelo marinho.

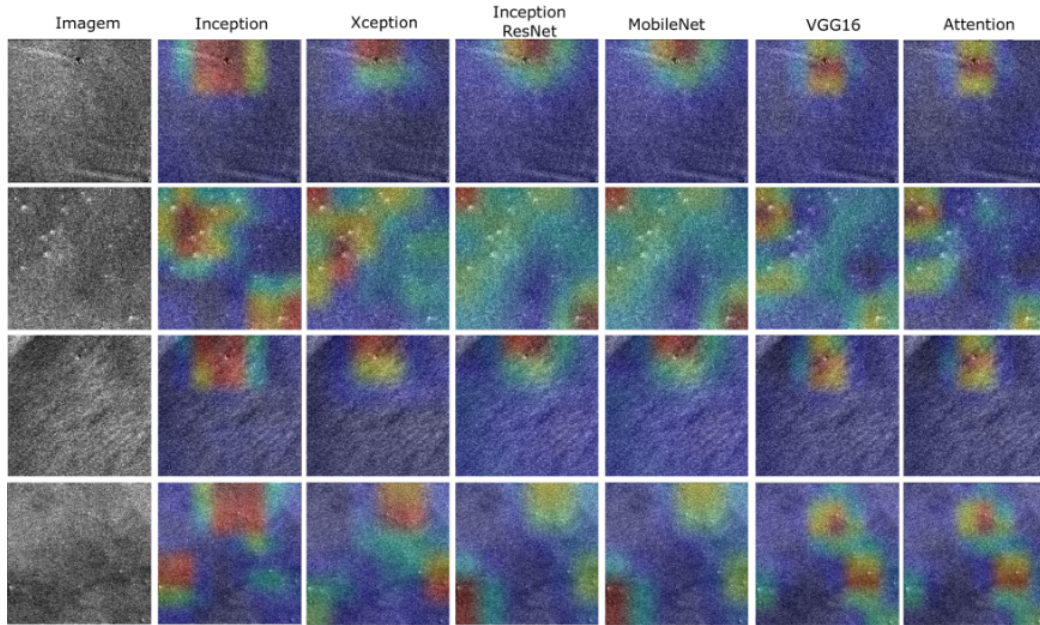


Figura B.7: GradCAM para amostras da classe Iceberg.

PUC-Rio - Certificação Digital Nº 1912831/CA

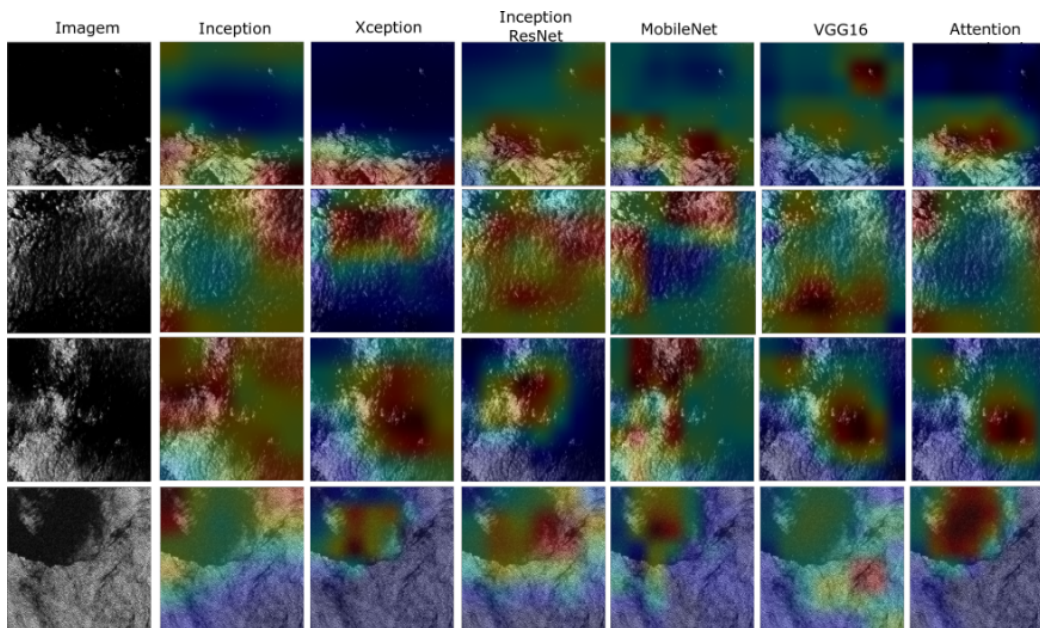


Figura B.8: GradCAM para amostras da classe Áreas de vento fraco.

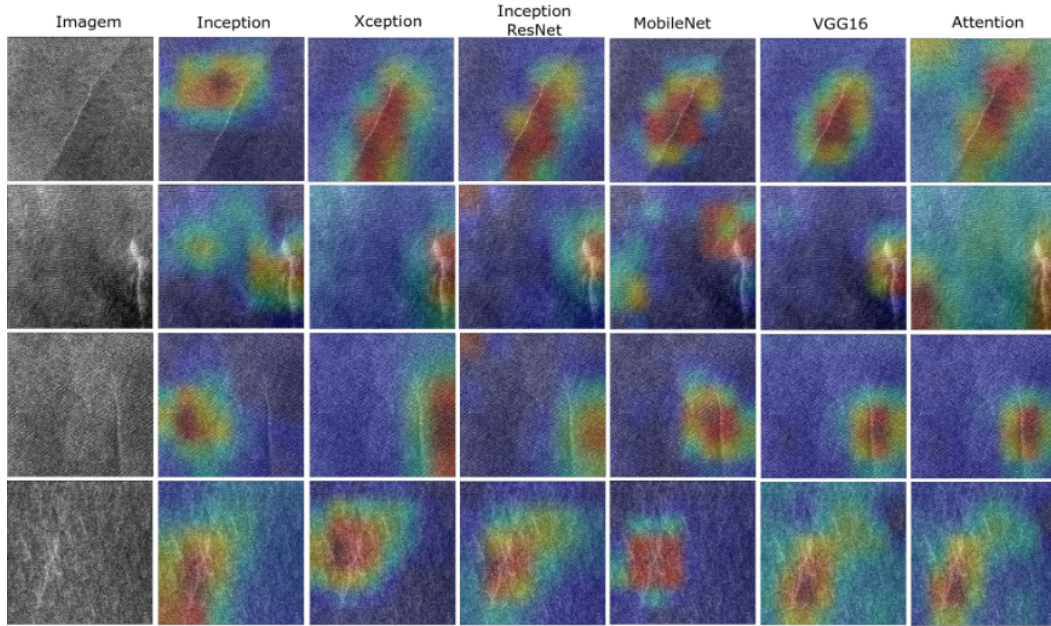


Figura B.9: GradCAM para amostras da classe Frente atmosférica.

PUC-Rio - Certificação Digital Nº 1912831/CA

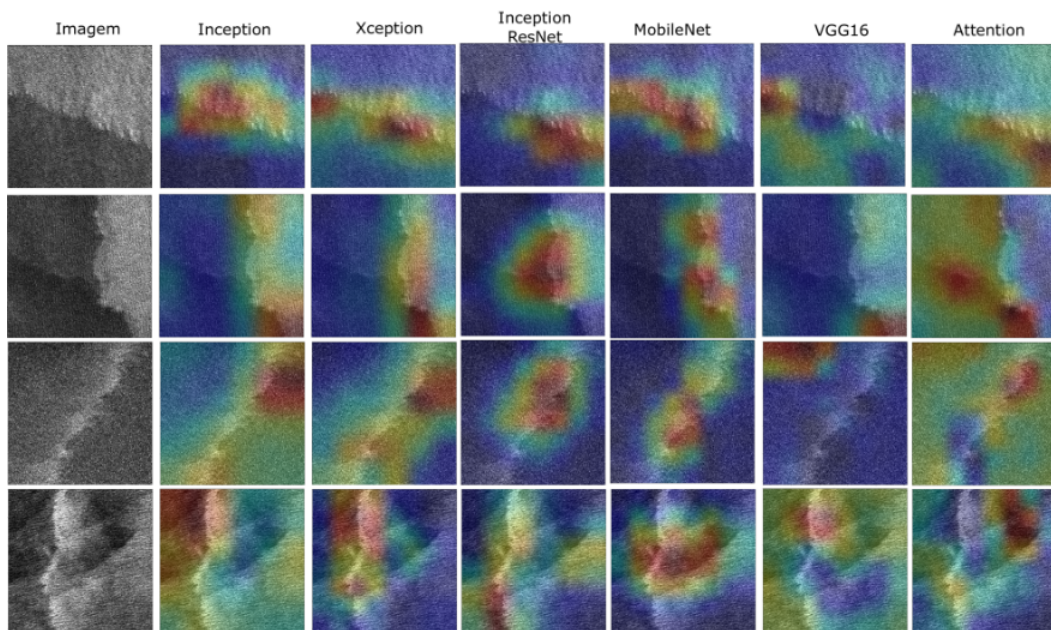


Figura B.10: GradCAM para amostras da classe Frente oceânica.