UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DAINF - DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

PAULA GIOVANNA RODRIGUES
VICTOR ANTONIO MENUZZO

ESTUDO COMPARATIVO DE ABORDAGENS COMPUTACIONAIS PARA CLASSIFICAÇÃO DE IMAGENS DE SATÉLITE DA AMAZÔNIA

CURITIBA
2020

PAULA GIOVANNA RODRIGUES
VICTOR ANTONIO MENUZZO

ESTUDO COMPARATIVO DE ABORDAGENS COMPUTACIONAIS PARA CLASSIFICAÇÃO DE IMAGENS DE SATÉLITE DA AMAZÔNIA

Proposta de monografia apresentada na matéria de Trabalho de Conclusão de Curso 1 da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do grau de bacharel em Sistemas de Informação

Orientadores: Leyza Baldo Dorini e Rodrigo Minetto
DAINF - Departamento Acadêmico de Informática - UTFPR

CURITIBA
2020

PAULA GIOVANNA RODRIGUES
VICTOR ANTONIO MENUZZO

ESTUDO COMPARATIVO DE ABORDAGENS COMPUTACIONAIS PARA CLASSIFICAÇÃO DE IMAGENS DE SATÉLITE DA AMAZÔNIA

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Sistemas de Informação da Universidade Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 04/Dezembro/2020

_____

RICARDO DUTRA DA SILVA
Doutorado
Universidade Tecnológica Federal do Paraná (UTFPR)

_____

JORGE LUIZ DOS SANTOS RAMOS JUNIOR
Graduação
Universidade Federal do Paraná (UFPR)

_____

LEYZA ELMERI BALDO DORINI
Doutorado
Universidade Tecnológica Federal do Paraná (UTFPR)

CURITIBA

2020

AGRADECIMENTOS

Agradecemos aos nossos orientadores professora doutora Leyza Baldo Dorini e professor doutor Rodrigo Minetto por toda a orientação e todo o tempo dedicado ao nosso aprendizado e evolução como profissionais, além de todos os recursos computacionais que nos disponibilizaram para que pudéssemos realizar o trabalho.

RESUMO

RODRIGUES, Paula; MENUZZO, Victor. Estudo comparativo de abordagens computacionais para classificação de imagens de satélite da Amazônia. 2020. 43 f. – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2020.

O monitoramento de grandes áreas de floresta é consideratdo algo de suma importância mesmo sendo extremamente difícil de ser realizado. A Amazônia é considerada a maior floresta tropical do mundo, um monitoramento com a finalidade de mapear os seus padrões de uso da terra torna-se mais fácil quando pensado a partir de imagens de satélite. Partindo dessa ideia, as redes neurais de aprendizado profundo representam a melhor forma de classificar grandes volumes de imagens com alta precisão. O presente trabalho traz uma comparação entre técnicas de aprendizado profundo aplicadas a um conjunto de dados composto de mais de 100.000 imagens da Floresta Amazônica, usando como métricas de avaliação o Fbeta e o tempo obtido por cada técnica na classificação. As imagens foram rotuladas com uma dentre quatro classes referente à condição climática e com nenhuma ou mais classes dentre treze opções referentes aos padrões de uso da terra apresentados na imagem. A conclusão final foi a de que a rede VGG16 seria a escolha mais apropriada para a resolução do problema, uma vez que além de obter um Fbeta de 91,79% - ficando entre os três maiores dentre as técnicas testadas - classificou a base de teste - composta de 61191 imagens - em tempo relativamente baixo: 92,2 segundos.

Palavras-chave: Classificação de imagens da floresta amazônica. Imagens de satélite. Monitoramento espacial. Reconhecimento de padrões em imagens. Aprendizado profundo. Problemas multi classes. Redes neurais.

ABSTRACT

RODRIGUES, Paula; MENUZZO, Victor. Título em inglês. 2020. 43 f. – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2020.

Monitoring large areas of forest is extremely important even though it is very difficult. The Amazon rainforest is currently considered the largest tropical forest in the world, monitoring it with the purpose of mapping its land use patterns becomes easier when thought from satellite images. Based on this idea, deep learning neural networks represent the best way to classify large volumes of images with high precision. The present work provides a comparison among deep learning techniques applied to a data set composed of more than 100,000 images of the Amazon rainforest, using Fbeta and the time obtained by each technique in the classification as the evaluation metrics. The images were labeled with one of the four classes related to the climatic condition and with none or more classes among the options related to the land use patterns. The final conclusion was that the VGG16 network would be the most appropriate choice for solving the problem, since in addition to obtaining an Fbeta of 91.79% - ranking top three among the tested techniques - it classified the test database - composed of 61191 images - in time relatively low: 92.2 seconds.

Keywords: Amazon rainforest image classification. Satellite imagery. Spatial monitoring. Image pattern recognition. Deep learning. Multi-class problem. Neural networks.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

O surgimento do monitoramento espacial via satélite possibilitou inúmeros estudos de acompanhamento de áreas de grande extensão e difícil acesso. Segundo Wulder e Coops (2014), imagens de satélite são a melhor fonte de dados para avaliações detalhadas do uso e cobertura da terra visando o monitoramento da perda de biodiversidade e da dinâmica do ecossistema.

Este trabalho tem como foco a análise de imagens de monitoramento espacial da Amazônia, considerada a maior floresta tropical do mundo. Ela é composta por uma área de 5.500.000 km^2 distribuída entre nove países, o que dificulta seu monitoramento por outros meios que não sejam via imagens de satélite. A importância de sua preservação é indiscutível. Por exemplo, ela serve como reguladora do clima, dado que pequenas mudanças em sua dinâmica já são capazes de afetar o CO_2 da atmosfera e influenciar na mudança climática (PHILLIPS et al., 2009).

Contudo, segundo o Instituto Nacional de Pesquisas Espaciais (INPE), órgão governamental que monitora a Amazônia Legal[1], houve um crescimento significativo do desmatamento a partir do ano de 2015, chegando a atingir uma área de 10.900 km^2 (veja o gráfico da Figura 1). Portanto, fica evidente a necessidade de métodos automatizados de análise de imagens de monitoramento espacial que permitam identificar eventos que ameacem a preservação ambiental, em especial o desmatamento.

Figura 1 – Evolução do desmatamento na Amazônia Legal - período entre os anos de 2008 e 2019.

Fonte: (ASSIS L et al., 2019)

[1]Parcela da Floresta Amazônica pertencente ao Brasil.

Os sistemas existentes para classificação de imagens de satélite da Amazônia - como o PRODES (INPE, 2019) - geram alertas sobre a ocorrência de desmatamento, mas sem especificar a atividade que o está ocasionando. Fearnside (1987) cita que os sistemas humanos de uso da terra são a raiz do padrão atual de desmatamento e que, portanto, devem ser o foco principal no estudo de políticas usadas para solucionar o problema. Desta forma, faz-se necessária uma abordagem capaz de monitorar os padrões de uso da terra ocorridos na Amazônia e não apenas o avanço do desmatamento na região.

Este trabalho propõe uma análise comparativa de abordagens de classificação baseadas em métodos de aprendizagem profunda, visando analisar imagens da Amazônia obtidas via satélite e classificá-las em diferentes fenômenos, os quais são de natureza climática e/ou de uso da terra.

Mais especificamente, foram realizados experimentos utilizando modelos clássicos de redes neurais convolucionais (VGG16, InceptionV3, ResNet50, MobileNet e MobileNetV2), bem como 12 combinações (*ensembles*) destes. A base de imagens utilizada foi disponibilizada pela empresa Planet via Kaggle e é composta de 101.670 imagens da Amazônia, obtidas entre 1º de janeiro de 2016 e 1º de fevereiro de 2017, sendo 40.479 rotuladas e 61.191 não rotuladas (PLANET, 2017).

Com base nos critérios considerados na análise (medida Fbeta resultante do teste cego do Kaggle e o tempo de classificação das imagens do conjunto de teste) a VGG16 foi considerada pela equipe a melhor opção de custo $\times$ benefício entre todos os experimentos realizados.

O restante do trabalho está organizado da seguinte forma. O Capítulo 2 detalha as abordagens utilizadas para resolução do problema e apresenta uma relação de trabalhos correlatos. O Capítulo 3 apresenta a metodologia empregada, e o Capítulo 4 realiza uma análise dos resultados obtidos pelas abordagens definidas anteriormente. O último capítulo, traz uma breve conclusão, juntamente com a discussão de trabalhos futuros.

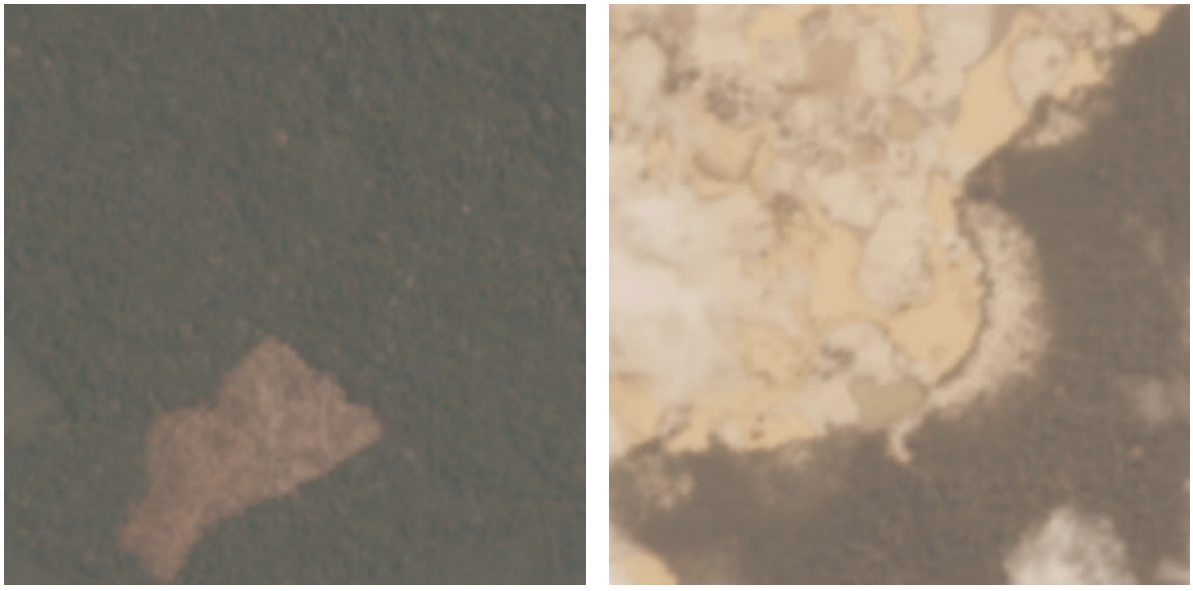
2 REVISÃO DE LITERATURA

A Seção 2.1 apresenta os principais aspectos relacionados à classificação em imagens de satélite. A Seção 2.2 discute brevemente abordagens que utilizam o mesmo conjunto de dados do presente trabalho. Por fim, a Seção 2.3 apresenta técnicas utilizadas para o desenvolvimento do trabalho.

2.1 Classificação em imagens de satélite

O monitoramento via satélite desempenha um papel importante na análise de informações geográficas, fornecendo amplas visões da superfície do solo e gerando um grande volume de dados (ABBURU; GOLLA, 2015). Yang e Lo (2002) e Sun et al. (2003) propõem a utilização de imagens de satélite para classificar fenômenos de uso da terra, ponto importante no combate ao desmatamento, conforme discutido no capítulo anterior. A Figura 2 ilustra exemplos deste tipo de imagem, onde ocorrem especificamente os fenômenos corte/queimada e mina artesanal.

Figura 2 – Imagens de satélite representando diferentes usos da terra.



(a) Corte/Queimada (b) Mina artesanal

Fonte: Planet (2017)

Para conseguir diferenciar determinadas classes de acordo com o conteúdo visual, é utilizado o processo da visão computacional definido como classificação (GONZALEZ; WOODS; EDDINS, 2004). Segundo Kaeli et al. (2015), este problema costuma ser um processo trivial para humanos, porém em termos computacionais ainda é um desafio.

Abburu e Golla (2015) realizaram pesquisas sobre métodos e técnicas de classificação para imagens de satélite. Os autores definem três métodos para realizar a classificação, os quais podem ser descritos como:

- Manual: são robustos e eficazes, porém demandam um elevado tempo para o processamento de grandes volumes de dados e exigem que os analistas estejam familiarizados com a área coberta pela imagem. Além disso, é um processo subjetivo, pois depende da interpretação e do conhecimento de cada analista (o que pode introduzir inconsistências na classificação).
- Automático: baseiam-se em algoritmos de classificação, sendo independentes de analistas e capazes de processar grandes volumes de dados.
- Híbrido: combina as vantagens dos métodos automatizado e manual. A abordagem utiliza classificação automatizada para fazer a classificação inicial e métodos manuais para refinar a classificação e corrigir erros.

No presente trabalho, serão considerados apenas métodos automáticos, os quais possibilitam propor abordagens de monitoramento em tempo real de grandes volumes de dados sem demandar a participação de analistas.

Yang, Everitt e Murden (2011) realizaram a classificação de plantações através de imagens de satélite utilizando métodos supervisionados (ou seja, aqueles que utilizam um conjunto de exemplos rotulados para aprendizado da regra, que será aplicada para mapeamento de novas entradas (DUDA; HART; STORK, 2001)). Foram testados cinco métodos de classificação: distância de Mahalanobis, distância euclidiana, *Spectral Angle Mapper* (SAM) e *Support Vector Machine* (SVM). Havia cinco classes no estudo: cultivos de milho, algodão, grãos, cana-de-açúcar e não-plantação. A utilização de SVM mostrou-se a técnica mais eficaz, uma vez que conseguiu obter precisão maior que 85% em todas as classes abordadas.

Akhtar, Nazir e Khan (2012) também estudaram o desenvolvimento da agricultura utilizando como base imagens de satélite. O principal diferencial em relação ao trabalho de Yang, Everitt e Murden (2011) foi em relação à extração de características das imagens. Para obtê-las foram aplicadas técnicas da Transformada Discreta de Cosseno (DCT) e da Transformada Discreta de Wavelet (DWT) para cada plantação. As imagens de satélite foram divididas em recortes com objetivo de identificar em cada um deles se havia ou não uma plantação. Com isso, as imagens puderam ser classificadas utilizando diferentes algoritmos (K-NN, SVM, Naive Bayes, SVM, Redes Neurais e Árvore de Decisão). Segundo os autores, devido a pequena quantidade de imagens, K-NN e Naive Bayes tiveram um desempenho melhor, chegando a obter aproximadamente 90% de acurácia.

Trabalhos mais recentes usam redes de aprendizado profundo, as quais são capazes de processar de forma mais eficaz um volume maior de imagens. A próxima seção apresenta alguns desses trabalhos, em específico aqueles que utilizaram a mesma base de imagens considerada no presente trabalho.

2.2 Trabalhos relacionados de aprendizado profundo

Os trabalhos a seguir são baseados em abordagens de aprendizado profundo e utilizaram para testes a mesma base de imagens considerada neste trabalho, a qual foi disponibilizada pela empresa Planet via Kaggle (PLANET, 2017). Detalhes específicos de algumas redes serão abordados na próxima seção.

Shi (2017), utilizou XGBoost e arquiteturas de redes convolucionais próprias para identificar as diferentes classes das imagens da Amazônia. Para realizar o XGBoost foram extraídas diferentes características, incluindo textura, cor e *keypoints* SIFT. Entretanto, o modelo não obteve bons resultados. Já a arquitetura convolucional era composta por duas camadas de convolução, uma de *pooling* e outra totalmente conectada. O resultado dessa arquitetura foi consideravelmente melhor, chegando a alcançar um índice Fbeta de 0.900. Apesar disso, nenhuma outra técnica como as combinações e/ou *transfer-learning* foi aplicada para aperfeiçoar o modelo.

Gardner e Nichols (2017), além de treinarem cada modelo separadamente, utilizaram *ensembles* para obter melhores resultados. Quatro arquiteturas de redes neurais convolucionais foram utilizadas: uma própria e três estado-da-arte (InceptionV3, ResNet50 e VGG16). O modelo que foi desenvolvido pelos autores obteve os piores resultados, pois só conseguia identificar classes que apareciam em quase todas as imagens. As três redes pré-treinadas obtiveram resultados similares e geraram o melhor resultado geral quando foram utilizadas juntas. Porém, não foram descritos detalhes dos critérios utilizados para a combinação.

Autores como Koguchi et al. (2018) encontraram através do aprendizado por transferência formas para aprimorar os resultados dos modelos. Para isso, treinaram duas arquiteturas de redes convolucionais, a VGG-16 e a Inception. Os pesos das redes do projeto de Shi (2017) foram utilizados para realizar o processo de aprendizado por transferência. Com isso, Koguchi et al. (2018) conseguiram aprimorar o índice Fbeta em 0.042 (indo de 0.877 para 0.919).

Um grande diferencial do projeto de Shendryk et al. (2018) foi demonstrar a capacidade de generalização, algo que não havia sido feito em nenhum dos trabalhos anteriores. As redes convolucionais foram treinadas utilizando as imagens da Amazônia e testadas em imagens nos trópicos úmidos da Austrália. O modelo utiliza redes similares àquelas apresentas anteriormente (InceptionV3, ResNet50 e VGG16) e obteve Fbeta de 0.91. A avaliação dos resultados nas imagens da Austrália foi feita apenas de forma visual, pois não foi obtido nenhum conjunto de dados para realizar a análise quantitativa.

Neste trabalho é proposta uma análise comparativa para a classificação de imagens de satélite utilizando abordagens de aprendizado profundo, as quais são discutidas na próxima seção.

2.3 Redes de aprendizado profundo

Técnicas de aprendizado profundo (LECUN; BENGIO; HINTON, 2015) têm sido exploradas em diversas aplicações, incluindo a classificação de imagens de sensoriamento remoto, foco deste trabalho. A ideia consiste basicamente em utilizar um grafo, composto por várias camadas de processamento, para modelar abstrações de alto nível nos dados.

Redes neurais convolucionais (do inglês, *Convolutional Neural Networks* - CNN) são provavelmente o modelo de aprendizado profundo mais famoso. Embora a ideia não seja nova — é creditada a Fukushima (FUKUSHIMA, 1980) a sua inspiração e a LeCun (Le Cun et al., 1989) diversas inovações importantes — foi em 2012 que Krizhevsky projetou uma rede de aprendizado profundo muito famosa, denominada AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), a qual venceu a competição ImageNet (RUSSAKOVSKY et al., 2015) (muito conhecida por promover grandes avanços na área de aprendizado de máquina) com uma ampla margem de ganho sobre o segundo colocado. Desde então, grandes avanços têm sido descritos, o que explica o seu domínio sobre algoritmos clássicos de aprendizagem de máquina como aqueles citados na Seção 2.1.

Redes convolucionais têm como princípio explorar propriedades topológicas que imagens, vídeos ou sinais de áudios possuem, ou seja, a relação que valores vizinhos compartilham (GOODFELLOW et al., 2016). Parâmetros que influenciam diretamente no comportamento e performance de uma rede convolucional são: número de filtros (*kernel*), tamanho dos filtros, *stride* (deslocamento do filtro ao longo da imagem de entrada), *pooling* (redução da dimensão espacial dos mapas ao longo das camadas da rede), camadas densas, escolhas de função de ativação, dentre outros tantos conceitos. Caso o leitor não esteja familiarizado com estes conceitos, sugere-se a leitura da bibliografia relacionada (LECUN; BENGIO; HINTON, 2015) (GOODFELLOW et al., 2016).

2.3.1 VGG16 e VGG19

Tanto a VGG16 quanto a VGG19 foram apresentadas no mesmo artigo por Simonyan e Zisserman (2015). Com a comparação das primeiras redes convolucionais – AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) e LeNet (Le Cun et al., 1989) – foi observado que a adição de mais camadas na rede levava a melhores resultados. Dado que adicionar mais camadas densas seria muito custoso computacionalmente, os autores investiram nas camadas de convolução. Para não ter que definir cada uma das camadas individualmente, foram criados blocos de convolução que se repetiriam ao longo da rede.

Ambas as redes iniciam com cinco blocos de convoluções ligados uns aos outros pela operação de *max pooling*, seguidos por três camadas densas e finalizando com uma função *soft-max*. Suas diferenças estão apenas nos últimos três blocos de convolução: a VGG19 apresenta quatro operações em cada bloco e a VGG16 apenas três. A Tabela 1 representa as arquiteturas das redes com mais detalhes.

Tabela 1 – Arquiteturas VGGs.

VGG16	VGG19
input	
conv 3x3x64	conv 3x3x64
conv 3x3x64	conv 3x3x64
max pooling	
conv 3x3x128	conv 3x3x128
conv 3x3x128	conv 3x3x128
max pooling	
conv 3x3x256	conv 3x3x256
conv 3x3x256	conv 3x3x256
conv 3x3x256	conv 3x3x256
	conv 3x3x256
max pooling	
conv 3x3x512	conv 3x3x512
conv 3x3x512	conv 3x3x512
conv 3x3x512	conv 3x3x512
	conv 3x3x512
max pooling	
conv 3x3x512	conv 3x3x512
conv 3x3x512	conv 3x3x512
conv 3x3x512	conv 3x3x512
	conv 3x3x512
max pooling	
FC-4096	
FC-4096	
FC-1000	
soft-max	

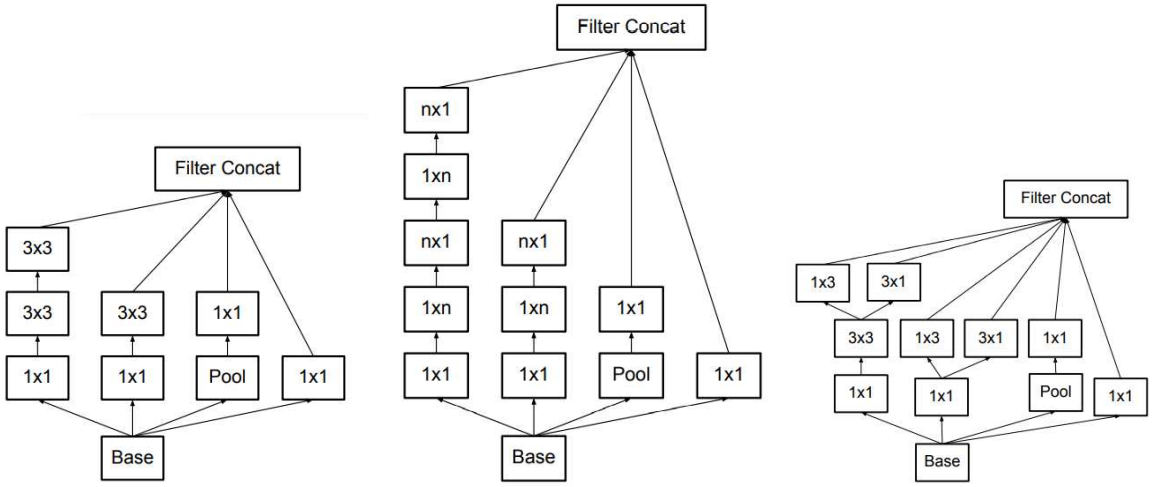
Fonte: Autoria própria

2.3.2 InceptionV3

Primeiramente definida por Szegedy et al. (2014) a arquitetura da Inception passou por algumas revisões. Mais especificamente, foi aprimorada em Szegedy et al. (2015), levando às redes InceptionV2 e InceptionV3, com a proposição de melhorias referentes ao desempenho computacional de redes de aprendizado profundo a partir da redução de gargalos entre as camadas.

Tais arquiteturas apresentam uma nova maneira de construção da rede, na qual camadas convolucionais são quebradas em outras menores (denominadas 'camada inception') e rodam em paralelo. Essa quebra diminui o número de parâmetros a serem transmitidos de uma camada para outra, conduzindo a uma maior eficiência computacional. A Figura 3 apresenta as três 'camadas inception' usadas pela InceptionV3 e a Tabela 2 detalha a arquitetura desse modelo.

Figura 3 – Camadas Inception.



(a) 1ª camada inception (b) 2ª camada inception (c) 3ª camada inception

Fonte: Szegedy et al. (2015)

Tabela 2 – Arquitetura InceptionV3

input	
conv 3x3	
conv 3x3	
conv 3x3	
pool	
conv 3x3	
conv 3x3	
conv 3x3	
1ª camada inception	*3X*
2ª camada inception	*5X*
3ª camada inception	*2X*
pool	
soft-max	

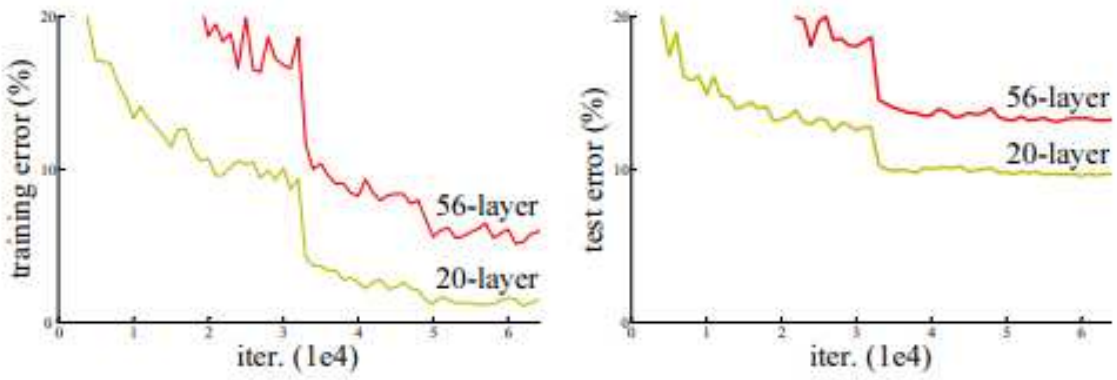
Fonte: Autoria própria

2.3.3 ResNet50

Desenvolvida por He et al. (2015), a ResNet50 foi criada a partir de um estudo que mostra que apenas o aumento das camadas de convolução nas redes de aprendizado profundo não garante a melhoria de performance. Diferentemente do que se pensava até então, os autores da ResNet perceberam que o aumento do número de camadas melhorava o desempenho da rede até certo ponto, a partir de um dado número de camadas adicionais a performance da rede piorava.A Figura 4 mostra o percentual de erro no treinamento e teste de duas redes diferentes no conjunto CIFAR-10 (KRIZHEVSKY, 2012), uma com 20

e outra com 56 camadas de convolução. Note que em ambos os cenários a rede com mais camadas tem maior percentual de erro.

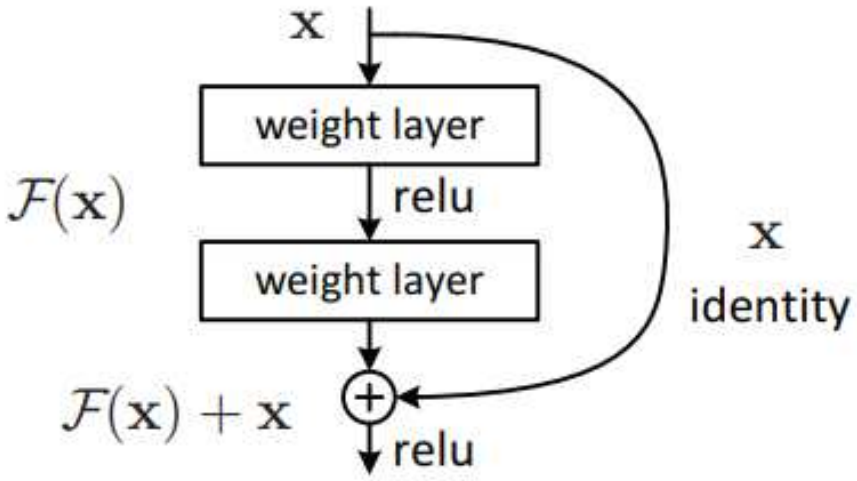
Figura 4 – Comparação da porcentagem de erro de treino e teste entre redes profundas com diferentes quantidades de camadas.



Fonte: (HE et al., 2015)

Para minimizar esse problema, a ResNet apresenta ligações entre camadas que não estão diretamente empilhadas. Essas ligações, ilustradas na Figura 5, servem para que a rede não perca resíduos importantes conforme se aprofunda. Os autores mostraram que o reaproveitamento desses resíduos traz ganhos significativos nas etapas de treinamento e testes das redes.

Figura 5 – Bloco de aprendizado residual da ResNet50.

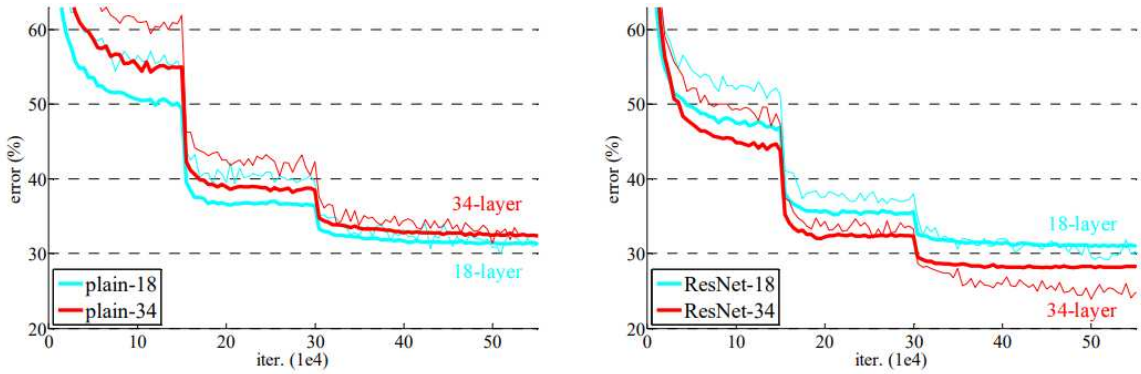


Fonte: (HE et al., 2015)

A Figura 6 mostra a comparação do percentual de erro entre duas redes sem aprendizado residual e duas redes com o aprendizado residual. Note que, quando o

aprendizado residual é inserido, a inclusão de mais camadas (ou seja, o aumento na profundidade da rede) conduz a melhorias nos resultados obtidos.

Figura 6 – Impacto no uso do bloco de aprendizado residual da ResNet50 ao utilizar diferentes quantidades de camadas.



Fonte: (HE et al., 2015)

A arquitetura da ResNet50 apresenta uma camada inicial de convolução ligada por um *max pooling* a quatro blocos de convolução, que se repetem 3, 4, 6 e 3 vezes e são ligados a uma camada totalmente conectada a partir de um *average pooling*. A rede é finalizada com a função *soft-max*. O aprendizado residual é feito como mostra a Figura 5 em todos os blocos de convolução. A Tabela 3 detalha melhor essa arquitetura.

Tabela 3 – Arquitetura ResNet50.

input	
conv 7x7x64	
max pooling	
conv 1x1x64 conv 3x3x64 conv 1x1x256	*3X*
conv 1x1x128 conv 3x3x128 conv 1x1x512	*4X*
conv 1x1x256 conv 3x3x256 conv 1x1x1024	*6X*
conv 1x1x512 conv 3x3x512 conv 1x1x2048	*3X*
avg pooling	
FC-1000	
soft-max	

Fonte: Autoria própria

2.3.4 MobileNet

Desenvolvidas para aplicativos móveis e de visão incorporada, as redes MobileNet são baseadas em arquiteturas simplificadas com o objetivo de apresentar redes profundas mais leves (HOWARD et al., 2017). Para melhorar a velocidade de processamento, essas redes empregam o conceito de convolução separada, ou seja, uma convolução para cada sinal de entrada da imagem (vermelho, verde e azul) seguida de uma convolução para combinar os resultados. Tal separação melhora o desempenho por facilitar os cálculos.

A Tabela 4 apresenta a arquitetura da MobileNet com maiores detalhes. As camadas que apresentam um *dw* são aquelas que se dividem em três convoluções para calcular os canais diferentes e voltam a se encontrar na camada seguinte.

Tabela 4 – Arquitetura MobileNet.

input	
conv 3x3x3x32	
conv 3x3x32 dw	
conv 1x1x32x64	
conv 3x3x64 dw	
conv 1x1x64x128	
conv 3x3x128 dw	
conv 1x1x128x128	
conv 3x3x128 dw	
conv 1x1x128x256	
conv 3x3x256 dw	
conv 1x1x256x256	
conv 3x3x256 dw	
conv 1x1x256x512	
conv 3x3x512 dw conv 1x1x512x512	*5X*
conv 3x3x512 dw	
conv 1x1x512x1024	
conv 3x3x1024 dw	
conv 1x1x1024x1024	
avg pooling	
FC - 1000	
soft-max	

Fonte: Autoria própria

A MobileNetV2 dá continuidade à proposta de ser uma rede de aprendizado profundo leve. Seguindo o mesmo modelo da MobileNet, sua contribuição está em adicionar camadas de residuais parecidas com as da ResNet, as quais otimizam o armazenamento de informações que podem se perder no processo.(SANDLER et al., 2018)

2.4 Combinações (*Ensembles*)

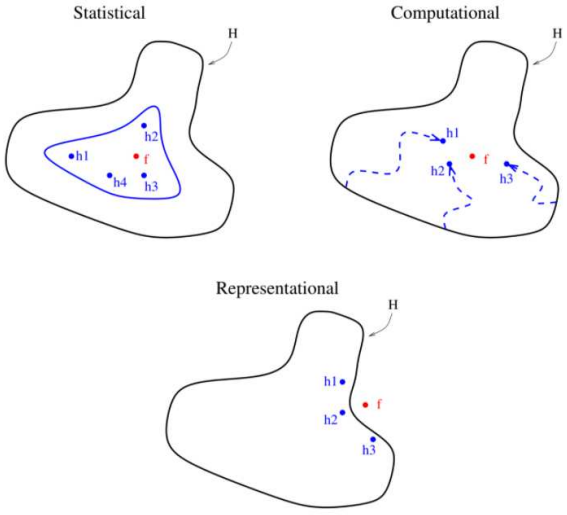
Como cada rede comete erros de generalização diferentes, uma decisão coletiva tem menor probabilidade de errar do que a decisão feita por qualquer uma das redes individualmente (Hansen; Salamon, 1990). Combinações são estratégias usadas para uma tomada de decisão coletiva, e se baseiam em uma votação entre duas ou mais redes de aprendizado diferentes para classificar determinado conjunto de dados.

Conforme descrito por Friedman et al. (2000), técnicas de combinações são utilizadas desde os primórdios da humanidade. Em seu texto, é exemplificado o caso de amigos que vão regularmente às pistas de corrida para apostar em cavalos e acabam fazendo uma descoberta: ao se considerar os dados de todos os apostadores, a chance de saírem vencedores nas apostas aumentaria.

Utilizar técnicas de combinação em problemas de *machine-learning* pode ser especialmente eficaz por três motivos principais (DIETTERICH, 2000), ilustrados na Figura 7:

- Estatístico: quando a quantidade de dados para treinamento é muito pequena, os algoritmos podem encontrar diversas hipóteses para resolução do problema. Realizar a média dessas hipóteses reduz o risco de escolher o classificador errado.
- Computacional: um classificador construído através de diferentes pontos de partida pode fornecer melhor aproximação da função a ser encontrada do que classificadores individuais.
- Representacional: quando a função desejada não pode ser representada, é possível expandir o espaço de representações utilizando somas ponderadas das hipóteses obtidas pelos classificadores.

Figura 7 – Motivos para utilização de combinações em *machine learning*.



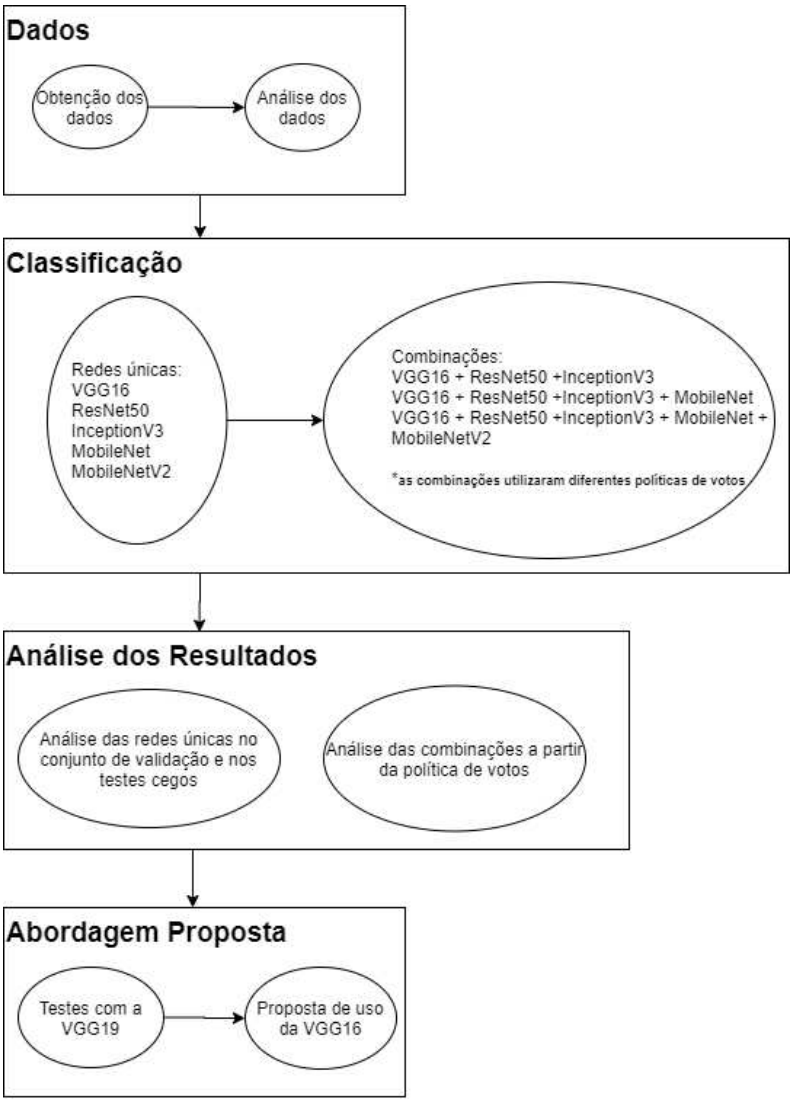
Fonte: (DIETTERICH, 2000)

3 MATERIAIS E MÉTODOS

Este capítulo tem como objetivo definir quais foram os dados e técnicas empregados para a classificação de imagens de monitoramento remoto da Amazônia. A Seção 3.1 descreve a base de imagens considerada e a Seção 3.2 discute os diferentes experimentos realizados.

A Figura 8 representa o fluxo de resolução do projeto. Inicialmente, foram feitas a coleta e a análise dos dados (Seção 4.1). Depois, foram definidas técnicas para resolução do problema e, a partir da avaliação dos resultados obtidos, foi escolhida a abordagem que apresentou o melhor custo vs. benefício.

Figura 8 – Fluxo de desenvolvimento do trabalho.



Fonte: Autoria Própria

3.1 Base de imagens de satélite

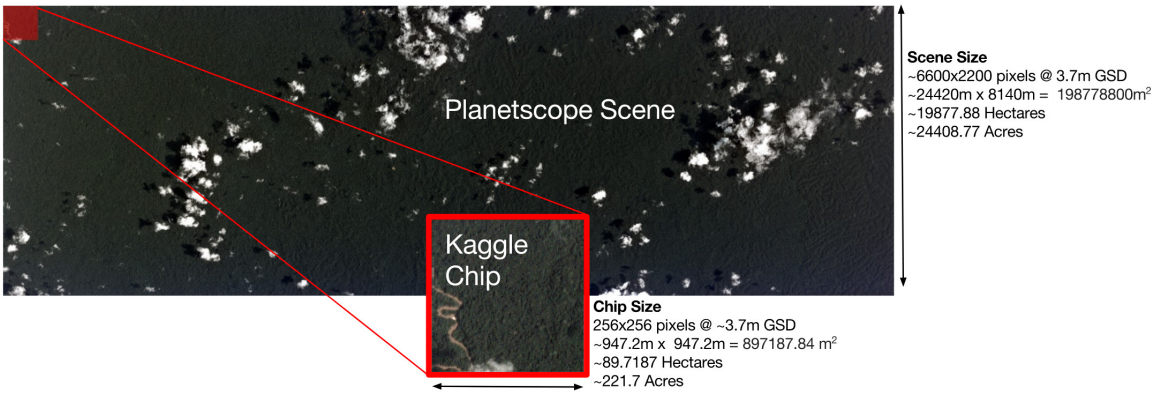
O presente trabalho utiliza imagens disponibilizadas pela empresa Planet via Kaggle derivadas de satélites de 4 bandas (*red, green, blue* e *near infra red*). A base é composta de 101.670 imagens da Amazônia obtidas entre 1º de janeiro de 2016 e 1º de fevereiro de 2017, sendo 40.479 rotuladas e 61.191 com rótulos conhecidos apenas pelo Kaggle (para posterior teste cego). Todas elas são da Bacia Amazônica que, como ilustrado na Figura 9, inclui ao todo noves países: Brasil, Peru, Colômbia, Venezuela, Guiana, Guiana Francesa, Suriname, Bolívia e Equador (PLANET, 2017).

Figura 9 – Bacia Amazônica.

Fonte: Geografia Visual

Como as dimensões das imagens de satélite são maiores que as desejadas para as redes, pois contemplam uma região muito extensa da Amazônia, elas estão divididas em recortes. Desta forma, é possível obter uma maior precisão quanto ao local de ocorrência de determinado fenômeno. A Figura 10 exemplifica a transformação feita de uma imagem de satélite para um recorte. Observe que, ao invés de manipular a imagem original de dimensão 6600×2200 pixels, representando 19877.88 hectares, é utilizado o recorte com dimensão 256×256 pixels, representando 89.7187 hectares.

Figura 10 – Exemplo de recorte (processo em que a imagem de satélite original é subdividida em regiões menores).

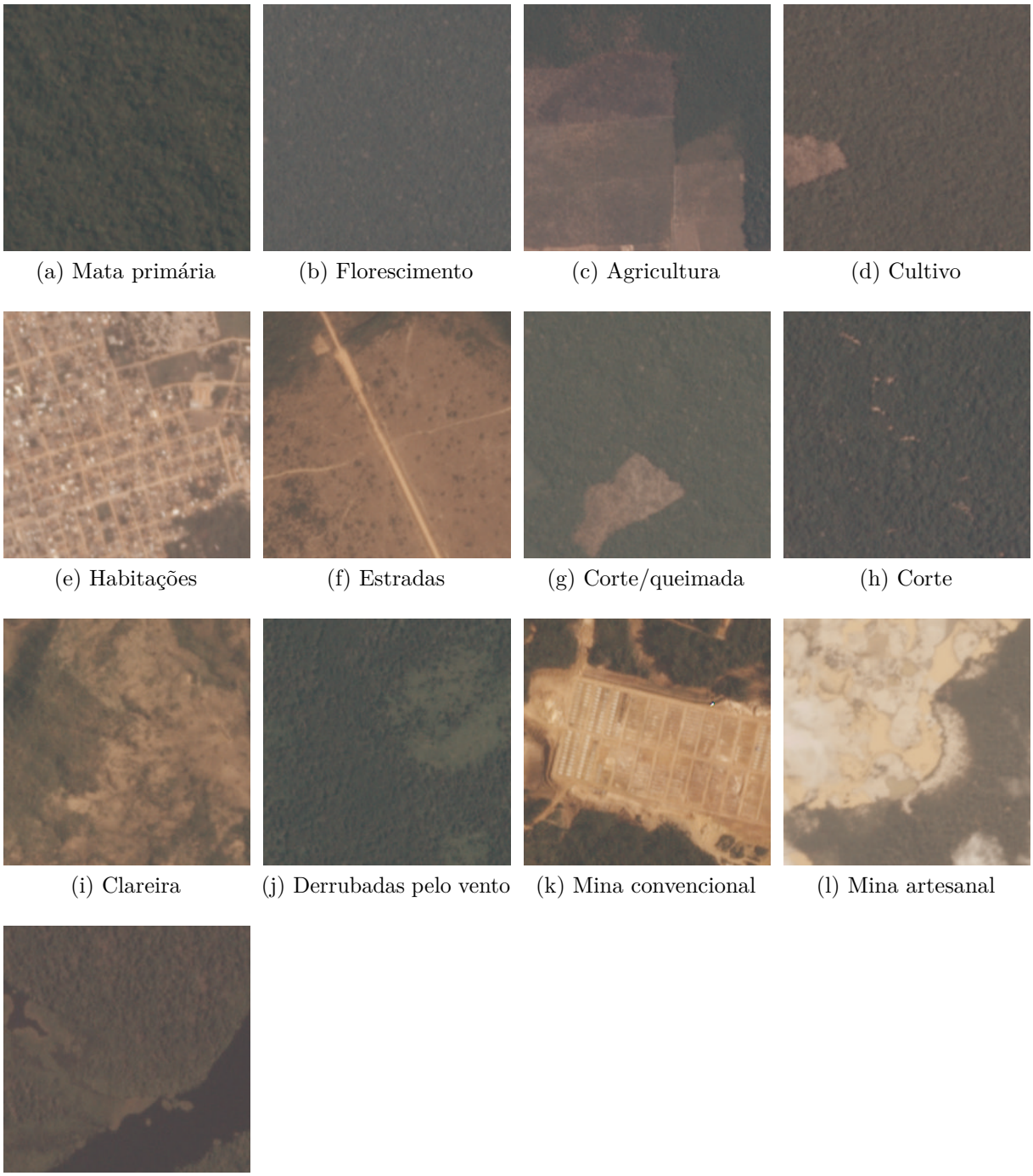


Fonte: Planet (2017)

As imagens podem ter 17 diferentes classificações, subdivididas em dois grupos principais: condições climáticas e uso da terra. A uma imagem é atribuído apenas um rótulo de condição climática, mas nela podem ocorrer diferentes fenômenos de uso da terra (posteriormente, serão discutidos aspectos mais específicos, como a ocorrência simultânea de diferentes classes em uma mesma amostra e o desbalanceamento da quantidade de ocorrências). As diferentes classificações são listadas e exemplificadas abaixo:

- Fenômenos de uso da terra, ilustrados na Figura 11:
  1. Mata primária;
  2. Florescimento;
  3. Agricultura - imagens que contêm agricultura voltada para fins comerciais;
  4. Cultivo - imagens que contêm agricultura de subsistência;
  5. Habitações;
  6. Estradas;
  7. Corte/queimada;
  8. Corte - imagens que contêm corte de um volume de árvores em quantidade significativa para caracterizar extração (e não um corte eventual);
  9. Clareira - imagens que contêm áreas naturalmente sem árvores;
  10. Derrubadas pelo vento - evento que ocorre na Amazônia em que o vento derruba árvores mais altas e abre clareiras na floresta.
  11. Mina convencional;
  12. Mina artesanal - se difere da mina convencional por conta da demarcação não ser tão específica e possuir grande presença de água;
  13. Água - imagens que contêm lagos ou rios.

Figura 11 – Exemplos das classes.



(a) Mata primária (b) Florescimento (c) Agricultura (d) Cultivo

(e) Habitações (f) Estradas (g) Corte/queimada (h) Corte

(i) Clareira (j) Derrubadas pelo vento (k) Mina convencional (l) Mina artesanal

(m) Água

Fonte: Planet (2017)

- Condições Climáticas, ilustradas na Figura 12:
  1. Nublado;
  2. Parcialmente nublado;
  3. Neblina;
  4. Tempo limpo.

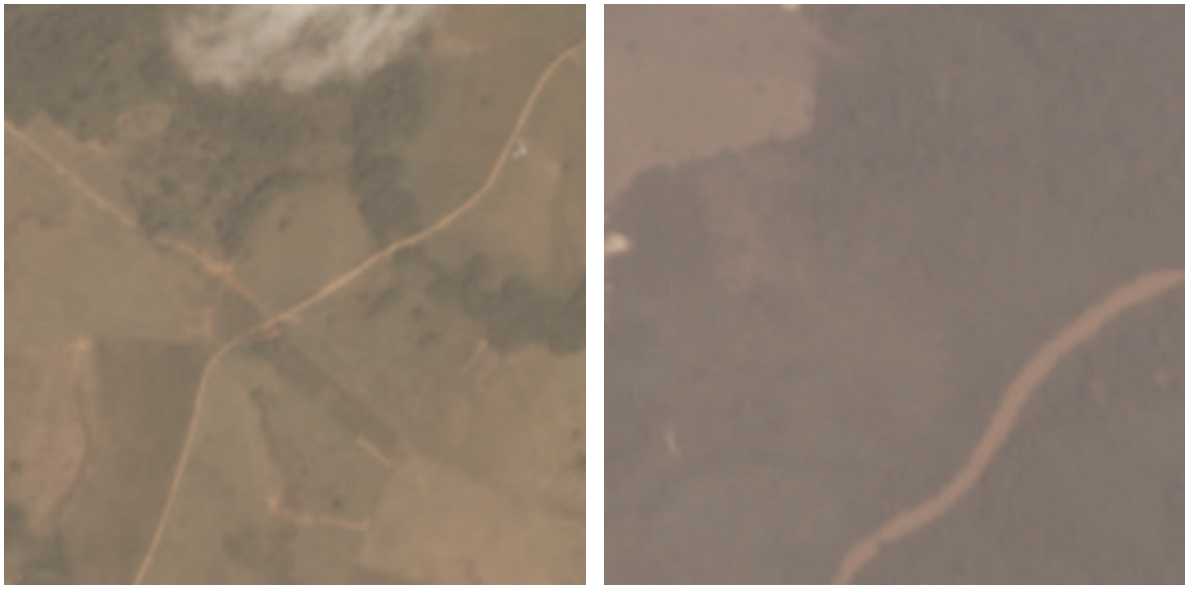
Figura 12 – Exemplos das classes relacionadas à condições climáticas.

(a) Nublado (b) Parc. nublado (c) Neblina (d) Tempo limpo

Fonte: Planet (2017)

As imagens também podem possuir casos de similaridade interclasse (ou seja, quando imagens de classes diferentes compartilham características similares) e dissimilaridade intra classe (ou seja, quando amostras da mesma classe possuem características bem distintas), aspectos desafiadores para o processo de classificação. A Figura 13 traz um exemplo de similaridade interclasse. Observe que a Figura 13(a), classificada como 'Estrada', se assemelha muito à Figura 13(b), classificada como 'Água'.

Figura 13 – Exemplo de similaridade interclasse.

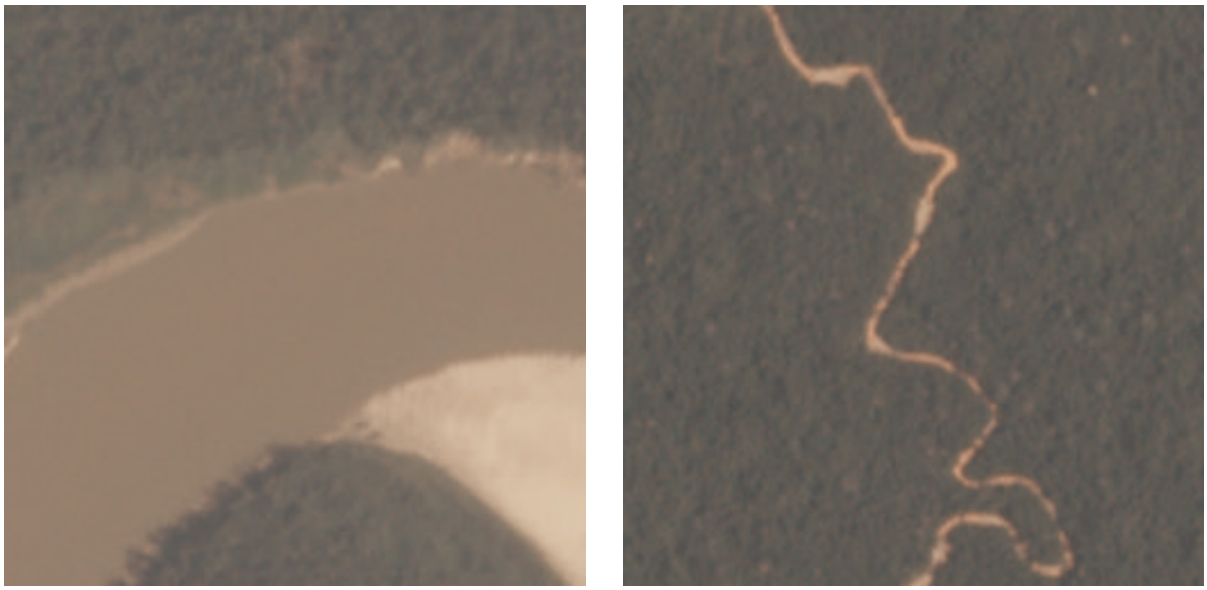


(a) Imagem classificada como Estrada (b) Imagem classificada como Água

Fonte: Planet (2017)

A Figura 14, por sua vez, traz um exemplo de dissimilaridade intraclasse presente na base. Ambas imagens pertencem à classe 'Água' apesar de possuírem algumas características diferentes. A Seção 4.1 irá apresentar análises mais aprofundadas referentes aos dados da base.

Figura 14 – Exemplo de dissimilaridade intraclasse.



(a) Exemplo 1 de imagem rotulada como Água (b) Exemplo 2 de imagem rotulada como Água

Fonte: Planet (2017)

O objetivo do trabalho consiste em identificar quais as classes presentes em cada uma das amostras. Para tal, foram realizados experimentos considerando diferentes abordagens de aprendizado profundo, conforme discutido na próxima seção.

3.2 Experimentos

Foram realizados testes considerando diferentes modelos clássicos de redes neurais convolucionais, bem como combinações destes. Mais especificamente, os seguintes experimentos foram realizados para classificação das imagens da base discutida na seção anterior:

- Utilização de uma única CNN: mais especificamente, foram realizados experimentos utilizando as redes VGG16, InceptionV3, ResNet50, MobileNet e MobileNetV2. A Seção 3.2.1 discute os parâmetros utilizados.
- *Ensembles*: dado que é possível que ocorram diferentes classes em uma mesma imagem, espera-se encontrar divergências entre as classificações. Portanto, é possível explorar a combinação dos resultados de diferentes redes. No total, foram consideradas 12 combinações, as quais são detalhadas na Seção 3.2.2.

O código foi desenvolvido em Python (ROSSUM; JR, 1995). Mais especificamente, as imagens foram pré-processadas utilizando a biblioteca OpenCV (BRADSKI, 2000) e as redes de aprendizagem profunda implementadas utilizando a biblioteca Keras (CHOLLET et al., 2015). Todos os modelos foram treinados e executados em uma máquina remota - composta por um processador Intel R CoreTM i7-3770K, GPU NVIDIA TITAN Xp e 32

GB RAM - sem a execução de processos paralelos.

3.2.1 Redes CNN

As redes VGG16, InceptionV3 e ResNet50 foram selecionadas com base no bom desempenho nos estudos de Gardner e Nichols (2017), Shendryk et al. (2018) e Koguchi et al. (2018) mencionados na Seção 2.2. Já as redes MobileNet e MobileNetV2 foram alternativas escolhidas pela equipe para a avaliação dos resultados de redes mais rápidas.

Para fins de análise, todas as redes foram treinadas com suas configurações padrão. Inicialmente, as imagens foram reduzidas de 256x256x3 para 128x128x3 - o canal infravermelho não foi utilizado -, o que contribuiu para melhoria do tempo de processamento. Visando evitar mínimos locais, a taxa de aprendizado utilizada foi otimizada pelo algoritmo Adam (KINGMA; BA, 2014). O critério de parada foi baseado na convergência da rede, ou seja, se o desempenho não melhorasse em mais de 10^{-4} durante cinco épocas consecutivas, o treinamento era finalizado.

3.2.2 Combinações

Visando minimizar os erros de generalização cometidos pelas redes individuais, foram realizadas combinações entre as redes utilizadas (como explicado na Subseção 2.4). Cada combinação consiste em três ou mais redes previamente treinadas e implementa uma política de votação, como definido na Tabela 5. A coluna 'Redes usadas' define quais redes foram selecionadas para formar cada combinação, enquanto a coluna 'Política de votos' define quantas redes do conjunto escolhido devem atribuir determinada classe à uma imagem para que a combinação rotule-a assim.

As políticas de votos foram um parâmetro fundamental no desempenho das combinações. Na Subseção 4.2.2 discute-se de forma mais aprofundada a interferência do grau de liberdade de cada uma das combinações, ou seja, qual foi o peso dessa política de votos na diferença entre o desempenho de combinações que continham o mesmo conjunto de redes usadas.

Tabela 5 – Combinações feitas.

Combinação	Redes usadas	Política de votos
01	VGG16+ResNet50+InceptionV3	>=1
02	VGG16+ResNet50+InceptionV3	>=2
03	VGG16+ResNet50+InceptionV3	>=3
04	VGG16+ResNet50+InceptionV3+MobileNet	>=1
05	VGG16+ResNet50+InceptionV3+MobileNet	>=2
06	VGG16+ResNet50+InceptionV3+MobileNet	>=3
07	VGG16+ResNet50+InceptionV3+MobileNet	>=4
08	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=1
09	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=2
10	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=3
11	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=4
12	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=5

3.3 Métricas de avaliação do modelo

O desempenho de cada experimento vai ser avaliado com base em duas métricas:

- Tempo: quantos segundos ele leva para classificar todas as imagens do conjunto de teste. Essa avaliação foi possível porque o modelo estava rodando em uma máquina remota sem nenhum outro processo concorrente que pudesse interferir no tempo de processamento.
- Medida Fbeta obtida a partir do teste cego das imagens classificadas. Ela corresponde à média harmônica entre os valores de precisão e revocação (SASAKI, 2007), conforme definido na Equação 1:

$$Fbeta^{1} = (1+\beta) \cdot \frac{precisão \cdot revocação}{(\beta^2 \cdot precisão) + revocação} \tag{1}$$

A precisão e a revocação são definidas nas Equações 2 e 3, respectivamente:

$$Precisão = \frac{VP}{VP + FP} \tag{2}$$

$$Revocação = \frac{VP}{VP + FN} \tag{3}$$

em que VP são os verdadeiros positivos, FP são os falsos positivos e FN são os falsos negativos.

[1] $\beta = 2$

4 AVALIAÇÃO DOS RESULTADOS OBTIDOS

A Seção 4.1 explora a análise das características do conjunto de dados, trazendo análises referentes à distribuição e ocorrência simultânea de classes, por exemplo. Posteriormente, na Seção 4.2 são discutidos os resultados obtidos.

Conforme apresentado no capítulo anterior, foram realizados experimentos considerando quatro modelos clássicos de redes neurais convolucionais (VGG16, InceptionV3, ResNet50, MobileNet e MobileNetV2), bem como 12 combinações (*ensembles*) destes. Como critérios de avaliação, foram utilizadas duas métricas: Fbeta calculada a partir da submissão cega dos resultados no Kaggle e o tempo que cada abordagem levou para classificar todas as 61191 imagens do conjunto de teste.

Por fim, com base na análise dos resultados obtidos, a Seção 4.3 apresenta uma proposta de abordagem baseada em técnicas de aprendizado profundo para classificação de imagens de satélite segundo as condições climáticas e de uso de terra previamente estabelecidas (Seção 3.1).

4.1 Análise dos dados utilizados

Esta seção discute análises referentes à distribuição e ocorrência simultânea de classes na base utilizada, visando identificar características de interesse no conjunto de dados. No conjunto de dados (PLANET, 2017), foram disponibilizadas 40479 imagens rotuladas, as quais foram separadas em conjuntos de treino e validação, sendo o último composto por 5479 imagens. Tais imagens foram utilizadas como base para este estudo. Além disso, cabe ressaltar que elas já correspondem aos recortes de dimensão 256×256 (veja discussão na Seção 3.1).

4.1.1 Distribuição das classes

A primeira análise realizada buscou identificar a quantidade de amostras disponíveis para cada classe. As Tabelas 6 e 7 deixam claro o desbalanceamento na distribuição de classes, tanto no conjunto de treinamento quanto no de validação.

Por exemplo, ao analisar a distribuição das classes relacionadas às condições climáticas (Tabela 6), observe que o rótulo ‘tempo limpo’ aparece em quantidade muito superior às demais.

Além disso, observe na Tabela 7 que, enquanto a classe referente a ‘mata primária’ está em aproximadamente 93% das imagens do conjunto de validação, as seis classes com menor representatividade - ‘Mina convencional’, ‘Derrubadas pelo vento’, ‘Corte/queimada’, ‘Florescimento’, ‘Corte’ e ‘Mina artesanal’ - somadas apresentam ocorrência menor que 4% do total de imagens.

Tabela 6 – Análise da distribuição das classes (condições climáticas).

Rótulo	Quantidade de ocorrências de cada classe	
	Conjunto de treinamento(35000 imagens)	Conjunto de validação (5479 imagens)
Nublado	1830 (5.23 %)	259 (4.72 %)
Neblina	2337 (6.68 %)	360 (6.57 %)
Parcialmente nublado	6266 (17.90 %)	995 (18.16 %)
Tempo limpo	24567 (70.19%)	3865 (70.54 %)
Total	35000	5479

Fonte: Autoria própria

Tabela 7 – Análise da distribuição das classes (condições de uso da terra).

Rótulo	Quantidade de ocorrências de cada classe	
	Conjunto de treinamento(35000 imagens)	Conjunto de validação (5479 imagens)
Mineração convencional	92 (0.26 %)	8 (0.15 %)
Derrubadas pelo vento	88 (0.25 %)	13 (0.24 %)
Corte e queimada	171 (0.49 %)	38 (0.69 %)
Florescimento	293 (0.84 %)	39 (0.71 %)
Corte	295 (0.84 %)	45 (0.82 %)
Mineração artesanal	290 (0.83 %)	49 (0.89 %)
Clareira	727 (2.08 %)	135 (2.46 %)
Habitações	3180 (9.09 %)	480 (8.76 %)
Cultivo	3924 (11.21%)	623 (11.37 %)
Água	6436 (18.39 %)	975 (17.80 %)
Estradas	6986 (19.96 %)	1085 (19.80 %)
Agricultura	10622 (30.35 %)	1693 (30.90 %)
Mata primária	32413 (92.61 %)	5100 (93.08 %)
Total	65517	10283

Fonte: Autoria própria

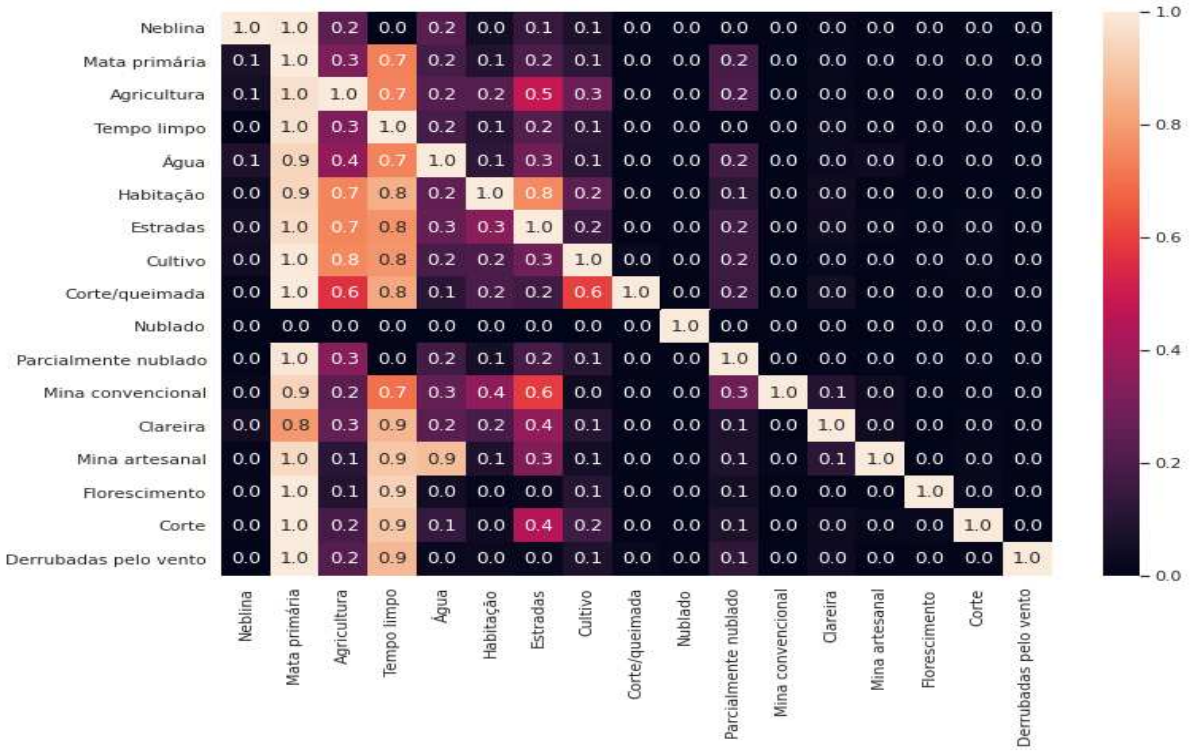
É importante lembrar que, apesar de uma imagem estar associada a apenas uma condição climática (‘tempo limpo’, ‘nublado’, ‘parcialmente nublado’ e ‘neblina’), é possível que nela ocorram diversas condições de uso de terra. Por exemplo, algumas amostras da base de treinamento podem chegar a possuir 14 diferentes classes. Por esta razão, o somatório das ocorrências de cada classe ultrapassa a quantidade de imagens.

4.1.2 Relações entre ocorrências de classes

Como em uma mesma imagem podem ocorrer diferentes classes, levantou-se a possibilidade de que a presença de determinada classe poderia implicar na presença de outra. Esse tipo de análise se torna especialmente importante quando as abordagens utilizadas falham ao reconhecer casos específicos.

O mapa da Figura 15 mostra a porcentagem de imagens rotuladas na classe A que também possuem a classe B. Por exemplo, quando ocorre a classe 'Mata primária', em 30% dos casos também estão presentes 'Agricultura' e em 70% 'tempo limpo'.

Figura 15 – Mapeamento das relações de ocorrência entre classes.



Fonte: Autoria própria

Pode ser observado a partir da leitura das linhas do gráfico, por exemplo, que 'Mata primária' e 'Tempo limpo' ocorrem juntamente com quase todas as demais classes (ou seja, quando uma amostra tem a classe 'Água' ela também possui 'Tempo limpo'). Por outro lado, as classes 'Nublado' e 'Corte' tendem a ocorrer sozinhas. Outros padrões a se destacar são:

1. relação entre a ocorrência de 'Corte/queimada' e 'Agricultura' (60%), bem como de 'Corte/queimada' e 'Cultivo' (60%);
2. nas imagens em que ocorre a classe 'Habitação', em geral também estão presentes 'Mata primária' (90 %), 'Agricultura' (70 %) e 'Estradas' (80 %);
3. a classe 'Estradas', que apresenta forte relação de ocorrência com 'Habitação' (80%), 'Mina convencional' (60%), 'Agricultura' (50%) e 'Corte'/'Clareira' (40%);

4. ao comparar condições de mineração, observa-se que, ao passo que 'água' ocorre em 90% dos casos para 'Mina artesanal', só está presente em 30% dos casos de 'Mina convencional'. Por outro lado, 'estradas' está em 60% das imagens de 'Mina convencional', mas em apenas 30% de 'Mina artesanal'.

Este mapeamento levantou algumas hipóteses sobre o quanto uma ou mais classes poderiam implicar no aparecimento de outra na mesma imagem. Esse tipo de análise se mostrou importante por conta do desbalanceamento da base, o que pode conduzir a uma maior dificuldade na classificação de classes com poucas amostras.

Para uma análise mais detalhada, foi criado um grafo com intuito de obter maiores informações sobre as ligações das classes que possuem poucas ocorrências (foi considerada apenas a base de treinamento, únicas amostras que possuem rótulos disponíveis). A Figura 16 ilustra o resultado obtido. Os nós representam as classes relacionadas às condições de uso da terra (excluindo 'mata primária', que está presente em praticamente todos os casos). Dois nós apresentam uma ligação cada vez que as duas classes aparecem juntas em uma mesma imagem. A espessura de cada ligação é correspondente ao peso dela, ou seja, classes que aparecem juntas frequentemente tem a espessura de sua ligação maior. Para facilitar a visualização dos pesos, eles também estão detalhados na Tabela 8.

Tabela 8 – Pesos das ligações (representando ocorrências simultâneas das classes) para o grafo ilustrado na Figura 16.

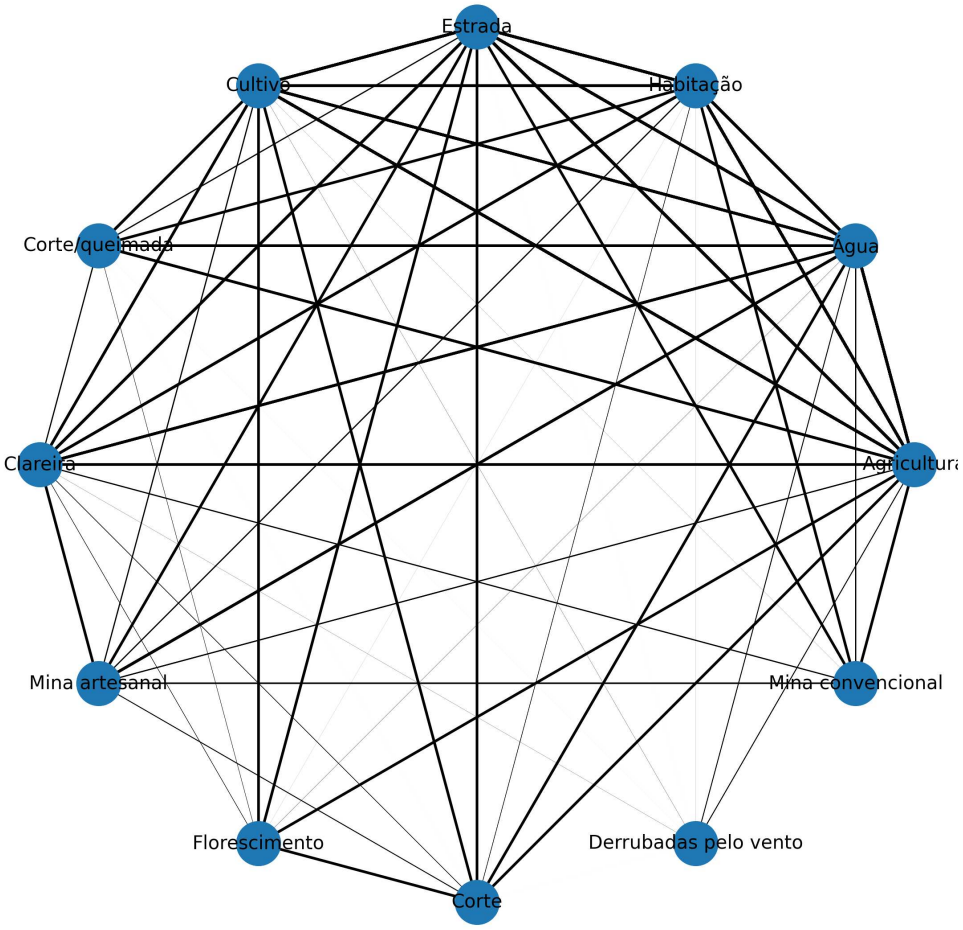
	01	02	03	04	05	06	07	08	09	10	11	12
Agricultura - 01	0	2712	2737	6034	3447	119	225	38	32	65	23	24
Água - 02	2712	0	915	2125	887	24	206	299	16	49	3	26
Habitação - 03	2737	915	0	2786	925	41	163	29	4	13	3	36
Estrada - 04	6034	2125	2786	0	1337	36	323	110	10	151	2	59
Cultivo - 05	3447	887	925	1337	70	130	89	18	35	58	8	4
Corte/queimada - 06	119	24	41	36	130	0	10	0	2	2	2	0
Clareira- 07	225	206	163	323	89	10	0	40	3	13	4	10
Mina artesanal - 08	38	299	29	110	18	0	40	0	0	6	0	4
Florescimento - 09	32	16	4	10	35	2	3	0	0	7	1	0
Corte - 10	65	49	13	151	58	2	13	6	7	0	1	0
Derrub. pelo vento - 11	23	3	3	2	8	2	4	0	1	1	3	0
Mina convencional - 12	24	26	36	59	4	0	10	4	0	0	0	0

Fonte: Autoria própria

Classes como 'Estradas', 'Água', 'Agricultura' e 'Cultivo' possuem muitas ligações fortes, ou seja, aparecem com frequência junto a outras. Isso pode ser explicado pelo fato que elas ocorrem com muita frequência no conjunto de dados, conforme destacado nas Tabelas 6 e 7.

Por outro lado, 'Mina artesanal', 'Corte/queimada', 'Florescimento' e 'Clareira' possuem um número menor de ligações fortes. Por serem pouco frequentes no conjunto

Figura 16 – Grafo de ligações entre as classes.



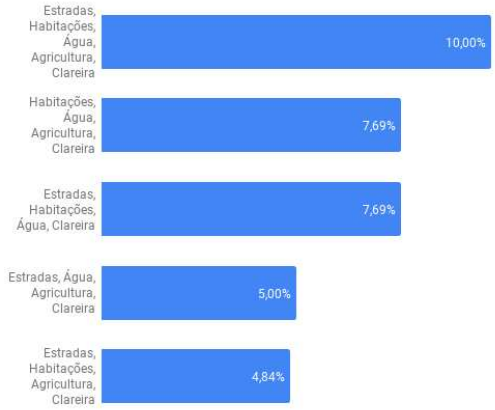
Fonte: Autoria própria

de dados, levanta-se à hipótese de que, na maioria dos casos em que uma delas aparece, pelo menos uma de suas ligações fortes tende a ocorrer também. Por fim, percebe-se que a classe 'Derrubada pelo vento' possui poucas conexões e as que existem são fracas (portanto, sua ocorrência não pode ser vinculada a nada).

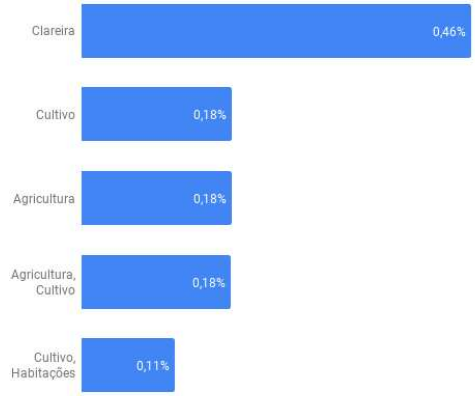
Com intuito de analisar se uma classe poderia ser induzida a partir da ocorrência de outras, foram analisadas as cinco ligações mais fortes para as seis classes com menor ocorrência na base de treinamento: 'Mina convencional', 'Derrubadas pelo vento', 'Corte/queimada', 'Florescimento', 'Corte' e 'Mina artesanal'. Como pode ser observado na Figura 17, essa relação não existe de forma significativa para os casos analisados.

Por exemplo, a classe 'Mina convencional' tem as ligações mais fortes com as classes 'Estrada', 'Habitação', 'Água', 'Agricultura' e 'Clareira'. Ao observar o conjunto de imagens que apresenta essas cinco classes juntas percebe-se que em 10% delas há também a classe 'Mina convencional'. A relação mais significativa entre a ocorrência de classes ocorre para 'Mina artesanal', a qual ocorre em 28.13% dos casos em que também estão presentes 'Água', 'Estradas' e 'Clareiras'.

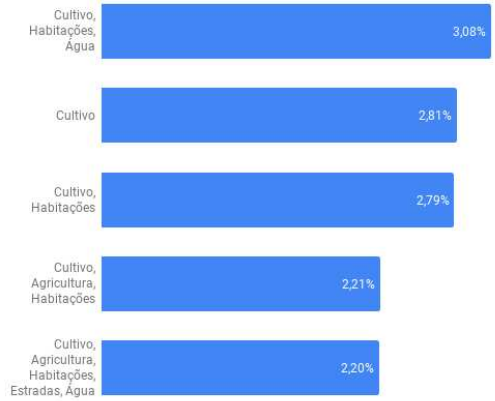
Figura 17 – Gráficos das cinco ligações mais significativas para as seis classes com menor ocorrência na base de treinamento.



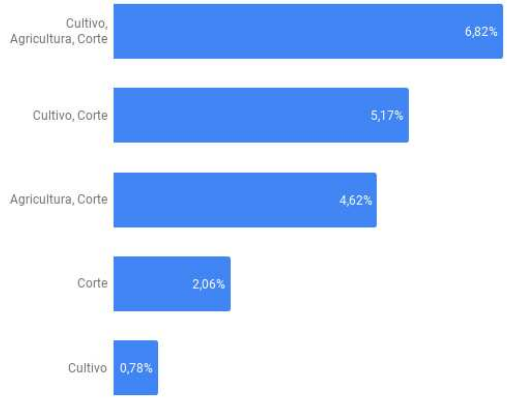
(a) TOP 5 Mina convencional



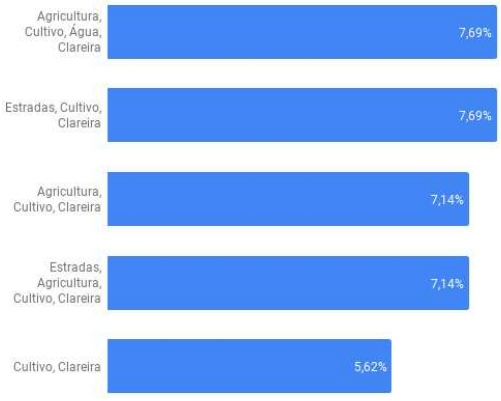
(b) TOP 5 Derrubadas pelo vento



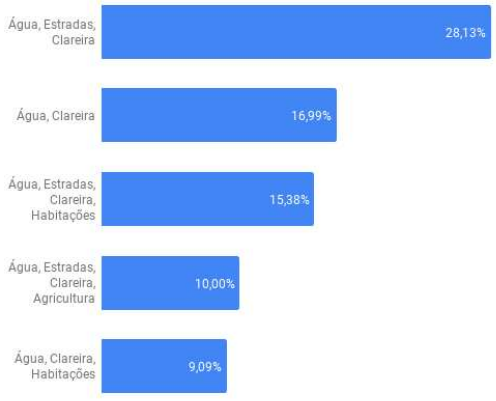
(c) TOP 5 Corte e queimada



(d) Top 5 Florescimento



(e) Top 5 Corte



(f) TOP 5 Mina artesanal

Fonte: Autoria própria

4.2 Resultados dos experimentos

Nesta seção, serão discutidos os resultados obtidos. Como detalhado no capítulo anterior, inicialmente foram realizados experimentos considerando quatro modelos clássicos de redes neurais convolucionais: VGG16, InceptionV3, ResNet50, MobileNet e MobileNetV2, cujos resultados são apresentados na Seção 4.2.1. Além disso, a Seção 4.2.2 traz a análise do desempenho de 12 diferentes combinações (*ensembles*) destas redes.

4.2.1 Resultados das redes isoladamente

A Tabela 9 mostra o percentual de acertos de cada rede em cada uma das classes no conjunto de validação. Ela está ordenada em ordem crescente conforme a quantidade de ocorrências das classes na base. Observe que a MobileNetV2 foi a única rede que teve problemas de classificação mesmo com as classes com mais ocorrências. Todas as demais tiveram porcentagens maiores que 50% nas dez classes com mais amostras.

Além disso, 'Mina convencional' é uma classe que acaba chamando atenção. Apesar de ser a classe com menos exemplares - não chegando a ter nem 0.15% de representatividade na base - acaba não sendo completamente ignorada por quatro das cinco redes analisadas. Uma hipótese que pode ser levantada ao observar o desempenho das redes em relação às classes de mineração - tanto 'Mina convencional' quanto 'Mina artesanal' - é que as correlações entre elas e as demais classes - como visto na Subseção 4.1 - provavelmente foram percebidas pela rede e auxiliaram no reconhecimento, mesmo com a baixa representatividade.

Tabela 9 – Porcentagem de acertos de cada rede no conjunto de validação.

Rótulo	VGG16	ResNet50	InceptionV3	MobileNet	MobileNetV2
Mina convencional	38%	12%	25%	12%	0%
Derrubadas pelo vento	0%	23%	0%	15%	0%
Corte e queimada	3%	3%	5%	5%	0%
Florescimento	38%	8%	26%	18%	0%
Corte	29%	33%	20%	13%	0%
Mina artesanal	69%	57%	59%	73%	6%
Clareira	36%	19%	24%	21%	2%
Nublado	94%	90%	91%	68%	85%
Neblina	78%	65%	79%	83%	11%
Habitações	79%	69%	71%	61%	5%
Cultivo	72%	52%	57.9%	56.9%	9%
Água	85%	77%	81%	86%	68%
Parcialmente nublado	97%	92%	89%	89%	46%
Estradas	86%	86%	85%	82%	37%
Agricultura	95%	82%	87%	87%	13%
Tempo limpo	98%	98%	98%	98%	98%
Mata primária	99%	99%	100%	100%	99%

A Tabela 10 apresenta a medida Fbeta obtida no teste cego do Kaggle para cada rede e o tempo - em segundos - que cada uma levou para classificar o conjunto de testes composto por 61191 imagens. Ressalta-se que o tempo foi medido com o modelo rodando em uma máquina remota sem nenhum outro processo concorrente que pudesse interferir no tempo de processamento.

Tabela 10 – Tabela de desempenho das redes nos testes cegos.

Rede	Fbeta	Seg.
VGG16	91,79%	97,21
ResNet50	89,89%	122,3
InceptionV3	89,94%	105
MobileNet	88,93%	74,6
MobileNetV2	72,75%	70,92

Fonte: Autoria própria

Percebe-se que o desempenho das redes no conjunto de validação é similar àquele do teste às cegas. Não foi possível fazer maiores análises referentes ao conjunto de testes porque os rótulos reais não foram disponibilizados pelo Kaggle. Entretanto, como o desempenho das redes ficou parecido com aquele obtido no conjunto de validação, pode-se concluir que os problemas enfrentados e os padrões encontrados foram os mesmos.

4.2.2 Resultados das combinações (*ensembles*)

Também foram realizados testes com as combinações das redes individuais utilizando um sistema de votação, visando compreender se a junção de três ou mais redes ajudaria a diminuir os erros das redes sozinhas. A Tabela 11 traz o resultado de todas as combinações testadas. Para facilitar a análise, essa tabela está ordenada em ordem decrescente em relação ao Fbeta.

Tabela 11 – Resultados das combinações feitas.

Comb.	Redes	Votos	Fbeta	Seg.
05	VGG16+ResNet50+InceptionV3+MobileNet	>=2	92,10%	300,2
09	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=2	91,80%	347,7
01	VGG16+ResNet50+InceptionV3	>=1	91,72%	222,7
02	VGG16+ResNet50+InceptionV3	>=2	91,63%	241,8
10	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=3	91,07%	364,9
04	VGG16+ResNet50+InceptionV3+MobileNet	>=1	90,90%	296,8
06	VGG16+ResNet50+InceptionV3+MobileNet	>=3	90,73%	290,9
11	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=4	88,10%	381,6
03	VGG16+ResNet50+InceptionV3	>=3	88,03%	325,6
08	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=1	86,73%	378,1
07	VGG16+ResNet50+InceptionV3+MobileNet	>=4	86,20%	282,1
12	VGG16+Resnet50+InceptionV3+MobileNet+MobileNetV2	>=5	73,96%	373,1

Fonte: Autoria própria

A primeira coisa que chama a atenção nos resultados apresentados na Tabela 11 é que o melhor Fbeta obtido entre as combinações ainda fica muito próximo ao Fbeta da VGG16 sozinha - uma diferença de 0.31 pontos percentuais. Isso indica que os padrões reconhecidos pelas redes para classificar cada uma das imagens foram muito parecidos.

Outro ponto a se destacar é a importância da interferência do grau de liberdade de cada uma das combinações: as tentativas que exigiam votação "Unânime" - 'Combinação 03', 'Combinação 07' e 'Combinação 12' - obtiveram as piores classificações. O Fbeta das três combinações foram inferiores ao de quatro das cinco redes testadas. Por outro lado, em combinações com um grau de liberdade extremamente alto, que exigiam apenas um voto, o Fbeta ficou muito dependente das redes escolhidas.

Além disso, note que na união das três melhores redes - 'Combinação 01' - os erros não tiveram peso suficiente para prejudicar de forma significativa o Fbeta, fazendo com que essa abordagem ficasse melhor classificada do que quatro das cinco redes testadas.

Já quando as redes com piores desempenhos no Fbeta no teste cego - 'MobileNetV2' e 'MobileNet' - entraram na combinação - 'Combinação 08' - os erros acabaram fazendo com que seu Fbeta ficasse pior do que o de quatro das cinco redes testadas.

4.3 Abordagem proposta

Por fim, com base nos resultados obtidos, foi proposta uma abordagem para classificação de imagens de satélite com base nas classes estabelecidas (condições climáticas e de uso da terra).

A Figura 18 representa a comparação de desempenho - levando em consideração o Fbeta resultante do teste cego do Kaggle e o tempo de classificação das imagens do conjunto de teste - entre todas as abordagens consideradas nos experimentos. A rede

VGG19 não entrou nas análises anteriores porque foi testada apenas no final do trabalho (por conta do bom desempenho da VGG16).

Figura 18 – Fbeta x Tempo de Classificação (Seg.) para todos os experimentos realizados (considerando redes únicas e *ensembles*)

Fonte: Autoria própria

A Tabela 12 apresenta o desempenho da VGG19 em cada uma das classes do conjunto de validação. Note que em apenas seis classes a VGG19 obteve resultado superior ao da VGG16 (‘Derrubadas pelo vento’, ‘Mina artesanal’, ‘Neblina’, ‘Cultivo’, ‘Estradas’, ‘Agricultura’). Seus resultados no teste cego também foram muito próximos: com Fbeta de 91.60%, a VGG19 ficou com apenas 0.19 pontos percentuais a menos do que a VGG16.

Tabela 12 – Porcentagem de acertos da VGG19 no conjunto de validação.

Rótulo	VGG19
Mina convencional	0%
Derrubadas pelo vento	8%
Corte e queimada	0%
Florescimento	13%
Corte	36%
Mina artesanal	78%
Clareira	10%
Nublado	96%
Neblina	86%
Habitações	68%
Cultivo	73%
Água	81%
Parcialmente nublado	94%
Estradas	89%
Agricultura	91%
Tempo limpo	98%
Mata primária	99%

Apesar de as combinações 05 e 09 terem tido melhores valores de Fbeta quando

comparadas àqueles da VGG16 - 92.10% e 91.8% respectivamente - seus tempos de classificação foram pelo menos três vezes maiores do que a rede sozinha.

Portanto, com base na discussão apresentada em todo este capítulo, a VGG16 é considerada pela equipe a melhor opção de custo × benefício entre as abordagens utilizadas.

5 CONSIDERAÇÕES FINAIS

O presente trabalho apresentou uma comparação entre técnicas de aprendizado profundo aplicadas ao problema de classificação de imagens da Bacia Amazônica em padrões de uso da terra e condições climáticas. Por se tratar de um problema multi classe, a avaliação do desempenho das técnicas com relação a determinadas classes foi dificultada, uma vez que o fato de uma imagem poder apresentar entre duas e quatorze classes não permitiu análises utilizando técnicas como matriz de confusão. Outro fator que dificultou um bom desempenho dos modelos foi a baixa representatividade de algumas classes no conjunto de imagens, o que fez com que algumas redes quase ignorassem a sua existência, resultando em um índice de acerto baixíssimo.

No geral, a rede que apresentou um melhor desempenho (com base na medida Fbeta e no tempo de classificação) foi a VGG16, com um Fbeta de 91,79% e um tempo de classificação de 91,2 segundos. As combinações 05 - VGG16, ResNet 50, InceptionV3 e MobileNet - e 09, - VGG16, ResNet 50, InceptionV3, MobileNet e MobileNetV2 - com uma política de votos que considerava as classes que obtinham dois ou mais votos, tiveram Fbeta superior ao da VGG16: de 92,1% e 91,8% respectivamente, mas seus tempos foram pelo menos três vezes maiores do que o da VGG16.

Para possíveis trabalhos futuros é interessante aprimorar os argumentos que foram utilizados para o treinamento das redes neurais convolucionais. Isso pode ser realizado utilizando uma busca por exaustão para obter as melhores combinações de configurações para cada rede. Essa alternativa pode resultar em significativa melhora do modelo em relação ao Fbeta. Utilizar outras formas para classificar as classes com pequena amostra e que consequentemente tiveram pior avaliação pelos modelos testados, é outro ponto que pode ser aprimorado.

Utilizar dados referentes a localização geográfica e temporais de cada imagem possivelmente aumentaria a precisão das classificações, uma vez que maiores correlações poderiam ser obtidas. Um estudo referente à implementação do modelo proposto em hardwares com menor poder de processamento seria relevante, possibilitando o uso da abordagem construída em drones e outros dispositivos similares (o que poderia ser aplicado no monitoramento de determinadas regiões da bacia amazônica em tempo real).

Referências

ABBURU, S.; GOLLA, S. B. Satellite image classification methods and techniques: A review. **International journal of computer applications**, Citeseer, v. 119, n. 8, 2015. Citado 2 vezes nas páginas 11 e 12.

Akhtar, A.; Nazir, M.; Khan, S. A. Crop classification using feature extraction from satellite imagery. In: **2012 15th International Multitopic Conference (INMIC)**. [S.l.: s.n.], 2012. p. 9–15. Citado na página 12.

ASSIS L, G. F. et al. Terrabrasilis: A spatial data analytics infrastructure for large-scale thematic mapping. **ISPRS International Journal of Geo-Information**, 2019. Citado na página 9.

BRADSKI, G. The OpenCV Library. **Dr. Dobb's Journal of Software Tools**, 2000. Citado na página 26.

CHOLLET, F. et al. **Keras**. GitHub, 2015. Disponível em: <https://github.com/fchollet/keras>. Citado na página 26.

DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. **International workshop on multiple classifier systems**. [S.l.], 2000. p. 1–15. Citado na página 20.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2. ed. [S.l.]: Wiley, 2001. ISBN 978-0-471-05669-0. Citado na página 12.

FEARNSIDE, P. M. Causes of deforestation in the brazilian amazon. **The geophysiology of Amazonia: vegetation and climate interactions**, Wiley New York, p. 37–61, 1987. Citado na página 10.

FRIEDMAN, J. et al. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). **The annals of statistics**, Institute of Mathematical Statistics, v. 28, n. 2, p. 337–407, 2000. Citado na página 20.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological Cybernetics**, v. 36, p. 193–202, 1980. Citado na página 14.

GARDNER, D.; NICHOLS, D. Multi-label classification of satellite images with deep learning. 2017. Citado 2 vezes nas páginas 13 e 27.

GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. **Digital image processing using MATLAB**. [S.l.]: Pearson Education India, 2004. Citado na página 11.

GOODFELLOW, I. et al. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado na página 14.

Hansen, L. K.; Salamon, P. Neural network ensembles. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 12, n. 10, p. 993–1001, 1990. Citado na página 20.

HE, K. et al. **Deep Residual Learning for Image Recognition**. 2015. Citado 3 vezes nas páginas 16, 17 e 18.

HOWARD, A. G. et al. **MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications**. 2017. Citado na página 19.

INPE. **PRODES**. 2019. Disponível em: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>. Citado na página 10.

KAELI, D. R. et al. **Heterogeneous Computing with OpenCL 2.0**. 1st. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2015. ISBN 0128014148, 9780128014141. Citado na página 11.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014. Citado na página 27.

KOGUCHI, C. et al. Understanding the amazon rainforest from space using neural networks. 2018. Citado 2 vezes nas páginas 13 e 27.

KRIZHEVSKY, A. Learning multiple layers of features from tiny images. **University of Toronto**, 05 2012. Citado na página 16.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105. Citado na página 14.

Le Cun, Y. et al. Handwritten digit recognition: applications of neural network chips and automatic learning. **IEEE Communications Magazine**, v. 27, n. 11, p. 41–46, 1989. Citado na página 14.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado na página 14.

PHILLIPS, O. et al. Drought sensitivity of the amazon rainforest. **Science (New York, N.Y.)**, v. 323, p. 1344–7, 04 2009. Citado na página 9.

PLANET. **Planet: Understanding the Amazon from Space**. [S.l.], 2017. <https://www.kaggle.com/c/planet-understanding-the-amazon-from-space>. Citado 9 vezes nas páginas 10, 11, 13, 22, 23, 24, 25, 26 e 29.

ROSSUM, G. V.; JR, F. L. D. **Python reference manual**. [S.l.]: Centrum voor Wiskunde en Informatica Amsterdam, 1995. Citado na página 26.

RUSSAKOVSKY, O. et al. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision (IJCV)**, v. 115, n. 3, p. 211–252, 2015. Citado na página 14.

SANDLER, M. et al. Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. **CoRR**, abs/1801.04381, 2018. Disponível em: <http://arxiv.org/abs/1801.04381>. Citado na página 19.

SASAKI, Y. The truth of the f-measure. **Teach Tutor Mater**, 01 2007. Citado na página 28.

SHENDRYK, I. et al. Deep learning-a new approach for multi-label scene classification in planetscope and sentinel-2 imagery. In: IEEE. **IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium**. [S.l.], 2018. p. 1116–1119. Citado 2 vezes nas páginas 13 e 27.

SHI, G. Use satellite data to track the human footprint in the amazon rainforest. In: . [S.l.: s.n.], 2017. Citado na página 13.

SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. 2015. Citado na página 14.

SUN, W. et al. Information fusion for rural land-use classification with high-resolution satellite imagery. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 41, n. 4, p. 883–890, 2003. Citado na página 11.

SZEGEDY, C. et al. Going deeper with convolutions. **CoRR**, abs/1409.4842, 2014. Disponível em: <http://arxiv.org/abs/1409.4842>. Citado na página 15.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. **CoRR**, abs/1512.00567, 2015. Disponível em: <http://arxiv.org/abs/1512.00567>. Citado 2 vezes nas páginas 15 e 16.

WULDER, M. A.; COOPS, N. C. Satellites: Make earth observations open access. **Nature News**, v. 513, n. 7516, p. 30, 2014. Citado na página 9.

YANG, C.; EVERITT, J.; MURDEN, D. Evaluating high resolution spot 5 satellite imagery for crop identification. **Computers and Electronics in Agriculture - COMPUT ELECTRON AGRIC**, v. 75, p. 347–354, 02 2011. Citado na página 12.

YANG, X.; LO, C. Using a time series of satellite imagery to detect land use and land cover changes in the atlanta, georgia metropolitan area. **International Journal of Remote Sensing**, Taylor & Francis, v. 23, n. 9, p. 1775–1798, 2002. Citado na página 11.