Classificando impacto humano sobre a floresta amazônica a partir de imagens de satélite

JONATHAN J. M. NUNES
Institute of Computing
University of Campinas
Campinas, Brazil
j146667@g.unicamp.br

FELIPE M. TAVARES
Institute of Computing
University of Campinas
Campinas, Brazil
f265680@g.unicamp.br

LUCAS DAVID
Institute of Computing
University of Campinas
Campinas, Brazil
lucas.david@ic.unicamp.br

RODRIGO A. M. FRANCO
Institute of Computing
University of Campinas
Campinas, Brazil
r233569@g.unicamp.br

***Resumo* – O monitoramento por satélite é um passo fundamental no combate à extração ilegal de árvores raras, à mineração irregular e à expansão desenfreada do agronegócio sobre áreas de preservação. Entretanto, esta não se configura como uma atividade trivial quando feita sobre a região da bacia do rio Amazonas. Com grandes áreas a serem cobertas e esforços contínuos de extratores aplicados de modo a esconder seu impacto sobre a região, a utilização de ferramentas de monitoramento autônomo se faz necessária. Neste trabalho, aplicamos técnicas baseadas em redes convolucionais e aprendizado profundo para solucionar o desafio denominado "Planet: Understanding the Amazon from Space". Testamos um conjunto de configurações estratégias distintas e reportamos seu efeito sobre o problema. Nossa melhor solução, baseada em *EfficientNet-B3* e no uso de *Stochastic Weight Averaging*, obtém um F_2*-score* de 90,14% no placar público e 89,90% no privado da competição.**

***Palavras-chave* – Desmatamento, monitoramento por satélite, redes convolucionais**

I. INTRODUÇÃO

A floresta amazônica compreende um dos maiores ecossistemas do mundo. Composto por grandes regiões de floresta tropical, ela abriga uma grande diversidade de espécies em ambos Reinos Animalia e Plantae. Acredita-se que a Amazônia contém mais da metade das espécies de plantas e animais no planeta, em sua grande maioria ainda não catalogados [1].

Florestas possuem um importante papel no ciclo de carbono de nosso planeta, absorvendo carbono e liberando oxigênio em curtos ciclos. A floresta Amazônica, especificamente, é responsável por produzir cerca de 6 porcento de todo o oxigênio no mundo [1]. Entretanto, quando árvores são cortadas e queimadas, quantidades de dióxido de carbono são liberadas de forma acelerada na atmosfera, contribuindo para mudanças graves (ou até mesmo perda) de *habitats*, mudança climática e vários efeitos devastadores decorrentes destes.

Consequências a longo prazo nesta região são também notáveis. Como o solo amazônico não é particularmente fértil, o ecossistema se sustenta por uma fina camada de material orgânico proveniente da própria floresta, formada a partir do depósito de restos de plantas e de material orgânico, denominada serrapilheira [2]. O *húmus* formado na serrapilheira, rico em nutrientes que alimentam as árvores, é manutenido por animais que habitam a região (predominantemente insetos), e protegido de insolação direta pela grande cobertura de folhas de árvores altas. Com a extração e queimada de uma área significativa, o húmus é destruído e o solo sedimentar é exposto, se mantendo assim por centenas de anos.

Naturalmente, o desflorestamento é uma atividade associada ao alto retorno financeiro. Árvores raras são seletivamente removidas, transportadas por rios ou estradas ilegalmente para o exterior e utilizadas na construção de itens (como móveis) de luxo, ou em projetos de arquitetura privados. A área de cobertura da floresta também é comumente reduzida por queimadas para a expansão desenfreada do agronegócio de larga escala (frequentemente assistida por autoridades ligadas ao agronegócio regional) ou de produções sub-existenciais de famílias ou comunidades isoladas. O desflorestamento (e outras formas severas de intervenção humana) amazônico, em especial, são recorrentes na bacia do rio amazônico. Compreendendo uma parcela expressiva da floresta, esta engloba países como Equador, Bolívia, Brasil e Peru [3]. Somente no Brasil, estima-se que 10851 km² de floresta foram desmatados em 2020, representando um crescimento de 7% em relação ao ano anterior [4].

II. A COMPETIÇÃO "PLANET: UNDERSTANDING THE AMAZON FROM SPACE"

Com a crescente preocupação sobre a pauta ecológica por parte da comunidade internacional, no início deste milênio, entidades governamentais e privadas foram formadas a fim de monitorar e punir desflorestamento descontrolado observado na Amazônia. Entretanto, esta não é uma tarefa trivial, considerando a grande extensão da floresta e práticas que buscam esconder a extração ilegal, como a extração em baixa escala e a extração seletiva.

De modo a impulsionar o desenvolvimento de técnicas e soluções de monitoramento autônomo, capaz de cobrir grandes áreas e indicar mais rapidamente (e consistentemente) os níveis de desmatamento, as empresas Planet e SCCON disponibilizaram em 2017 a competição "Planet: Understanding the Amazon from Space"na plataforma Kaggle, com um conjunto de dados de mesmo nome. A competição reuniu 936 grupos competidores entre 20 de abril e 20 de julho de 2017, e distribuiu premiações em dinheiro para os três primeiros

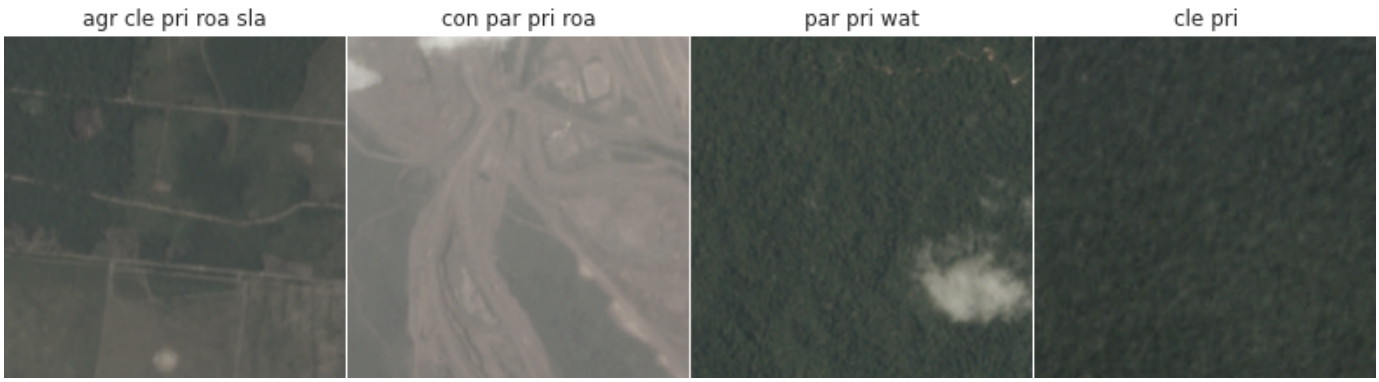

Figura 1. Exemplos de imagens no conjunto de dados "Planet: Understanding the Amazon from Space". As iniciais das *labels* associadas estão descritas acima de cada imagem.

colocados. Era esperado dos grupos o desenvolvimento de uma solução capaz de identificar os diversos tipos de intervenção humana sobre a região amazônica, representadas a partir de imagens aéreas feitas por satélite sobre a região (Fig. 1).

O conjunto de imagens, de mesmo nome que a competição, foi previamente subdividido em dois. O conjunto de treino, primeiramente, possui 40.479 amostras. Cada amostra é composta por uma imagem de tamanho 256 por 256 — uma captura feita por satélites de 4 bandas sobre uma região correspondente a uma área de $897.187,84$ m² — e um vetor de rótulos (ou *labels*) ocorrentes naquela imagem. As *labels* possivelmente associadas às amostras são, em ordem alfabética: *agriculture*, *artisinal mine*, *bare ground*, *blooming*, *blow down*, *clear*, *cloudy*, *conventional mine*, *cultivation*, *habitation*, *haze*, *partialy cloudy*, *primary*, *road*, *selective logging*, *slash burn* e *water*

O histograma na Fig. 2 apresenta a distribuição de rótulos do conjunto. É notável um alto desbalanceamento neste conjunto, onde (a) os rótulos *primary*, *clear* são predominantes; (b) *agriculture*, *road* e *water* são frequentes e *artisinal mine*, *blooming*, *blow down*, *selective logging* e *slash burn* são extremamente raros. Ademais, observa-se uma alta co-ocorrência

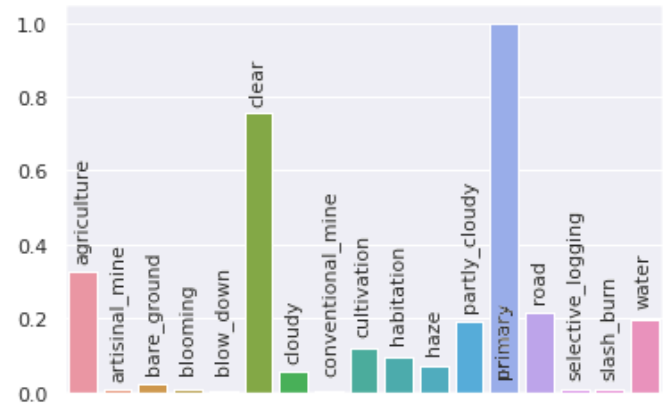

Figura 2. Histograma de *labels* no conjunto de dados Planet: Understanding the Amazon from Space. Nota-se que se trata de um problema extremamente desbalanceado, com a predominância das *labels agriculture*, *primary* e *clear*.

de sub-grupos de rótulos específicos (Fig. 3), configurando-se assim um problema com rótulos não regularmente distribuídos.

O conjunto de teste é formado por 61.191 imagens de satélite, similares às presentes no conjunto de treino. Os rótulos das imagens de teste não estão publicamente disponíveis, de modo a manter a validade do conjunto de teste para validações de modelos mesmo após o fim da competição.

O conjunto de teste da competição também foi subdivido em dois, o conjunto de teste público consistindo em aproximadamente 66% dos dados do conjunto de teste total, e o conjunto de teste privado consistindo em aproximadamente 34% dos dados do conjunto de teste total. Essa separação foi feita para permitir o acompanhamento do *leaderboard* pelo "*leaderboard* público" (usando dados do conjunto de teste público) durante a competição, e ao fim os resultados foram também disponibilizados no "leaderboard privado" (usando dados do conjunto de teste privado), qual refletiu as posições "finais" do desafio.

O objetivo da competição é decidir, dado cada amostra no conjunto de teste, quais rótulos estão associados a esta imagem. A avaliação dos resultados é feita internamente, pela própria plataforma Kaggle. A métrica utilizada na competição é a F_2-score [5], comumente empregada em problemas de classificação *multi-label* [6]. Formalmente, ela é definida por:

$$F_\beta \text{ score} := (1+\beta^2)\frac{\text{precision} \cdot \text{recall}}{\beta^2 \text{precision} + \text{recall}} \qquad (1)$$

onde $\beta = 2$.

Finalmente, observa-se pela pontuação da competição [1] que os três primeiros colocados foram o times "bestfitting", "Plant" e "Russian Bears", obtendo respectivamente as pontuações $93{,}317\%$, $93{,}294\%$ e $93{,}277\%$. Todos esses times utilizaram soluções baseadas em redes convolucionais profundas e técnicas de fortificação ou redundância estatística, como *ensemble* e *answer fusion*, e métodos clássicos de visão computacional para remoção de neblina. Estas serão detalhadas na Seção V.

[1] kaggle.com/c/planet-understanding-the-amazon-from-space/leaderboard

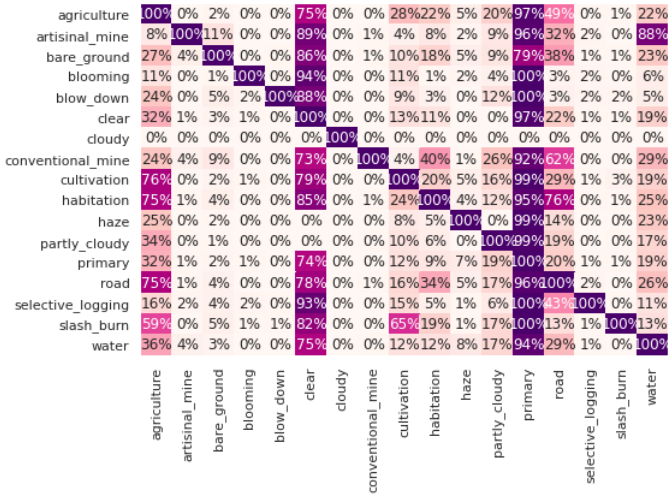

Figura 3. Co-ocorrência de *labels* no conjunto de dados "Planet: Understanding the Amazon from Space". Há uma grande correlação entre a ocorrência de diversos sub-grupos, como (a) *artisinal_mine* e *water*; (b) *habitation*, *agriculture*, *primary*, *clear* e *road*.

III. Disciplina Aprendizado de Máquina e Reconhecimento de Padrões

Este trabalho está sendo considerado como projeto final da disciplina MO444 — Aprendizado de Máquina e Reconhecimento de Padrões — ministrado pela professora Esther Colombini e PEDs (participantes do Programa de Estágio Docente) Alana Correia e Patrick Ferreira, do Instituto de Computação da Universidade Estadual de Campinas (Unicamp) no primeiro semestre de 2021. Cada membro do grupo contribuiu para o desenvolvimento deste trabalho de maneira colaborativa, seja na busca de assuntos e métodos para serem abordados na pesquisa como no desenvolvimento e implementação da solução (código a ser disponibilizado na entrega do trabalho). No entanto, é possível dizer que os membros também tiveram envolvimento mais especializado na condução das seguintes atividades:

- Jonathan J. M. Nunes: Exploração do método de *Focal Loss*.
- Felipe M. Tavares: Exploração do método de *Model Averaging* por *Stochastic Weight Averaging* e sua implementação a partir do *baseline*.
- Lucas David: Processamento dos dados, implementação do modelo *baseline*, exploração da estratégia de balanceamento de dados e do método de *Spatial Pyramid Pooling*.
- Rodrigo A. M. Franco: Exploração do método *Auto Augmentation* e otimizador AdamW.

Uma condição para escolha do tema do trabalho final da disciplina é a relação com ao menos um dos Objetivos de Desenvolvimento Sustentável no Brasil, comentado na próxima seção. Finalmente, Uma apresentação sobre nosso trabalho está disponível no YouTube[2].

[2]youtu.be/4ZMSX-1PhoM

IV. Objetivos de Desenvolvimento Sustentável no Brasil

Os Objetivos de Desenvolvimento Sustentável fazem parte de esforços de nível global para acabar com a pobreza, proteger o meio ambiente e o clima, e garantir que pessoas em qualquer lugar possam viverem em paz e prosperidade. Os objetivos estão descritos na página da Organização das Nações Unidas (ONU). Acordados com a ONU em 2015, 195 nações pretendem a partir dos objetivos estabelecidos melhorar a vida das pessoas em seus países até 2030.

O assunto escolhido e a pesquisa de técnicas de Visão Computacional e Aprendizado de Máquina para classificação de imagens de satélites da Floresta Amazônica estão ligados com o objetivo de desenvolvimento sustentável da agenda 2030 no Brasil "(13) Ação contra a mudança global do clima", mais especificamente ao item "13.1 Reforçar a resiliência e a capacidade de adaptação a riscos relacionados ao clima e às catástrofes naturais em todos os países", onde a contribuição para melhoria de técnicas para classificação de imagens de satélite na floresta amazônica, pode ajudar no monitoramento e detecção de atividades ilegais que podem vir a acelerar o processo de mudança climática.

V. Trabalhos relacionados

Nesta seção, detalhamos os estudos relacionados e conceitos utilizados em nosso trabalho.

A. Redes convolucionais

Na última década, redes convolucionais apresentaram grande poder de generalização sobre diversos contextos diferentes, sendo frequentemente empregadas na solução de problemas de classificação [7], detecção [8], segmentação [9], estimativa de pose [10] e diversos outros [11]. Trabalhos aplicados à imagens capturadas por ferramentas de monitoramento remoto (como satélites) também foram conduzidos recentemente de modo a identificar e demarcar automaticamente regiões de interesse [12].

Redes convolucionais consistem na aplicação sucessiva da operação de convolução, definido como a multiplicação no domínio de frequência entre o sinal espacial (uma imagem) e um *kernel* ou filtro. Adicionalmente, o sinal tem suas dimensões espaciais recorrentemente reduzidas com o emprego de camadas de *pooling* ou $stride > 1$, o que aplica a convolução deslizando a janela em deslocamentos maiores que 1.

B. Transfer learning

O aprendizado por transferência (ou *transfer learning*), no contexto de aprendizagem profunda, consiste na construção de representações discriminativas de padrões, muitas vezes em múltiplos níveis de complexidade, que podem ser generalizadas para múltiplos problemas [13]. Por promover uma redução significativa na quantidade necessária de dados para treino, tempo de treinamento e possivelmente melhorando resultados, transferência se mostrou extremamente popular nos últimos anos, em ambas comunidades científica e comercial.

Representações visuais generalizáveis podem ser obtidas a partir de treinamentos sobre grandes bases de dados, usualmente envolvendo múltiplos padrões, entidades e classes. *Frameworks* modernos muitas vezes promovem modelos pré-treinados sobre a base de dados ILSVRC [14].

C. Data Augmentation

Aumentação de dados, também chamado de "enriquecimento de dados", é utilizado como uma forma de melhorar a generalização de algoritmos, através da adição de novos exemplos ao conjunto de dados de forma artificial [15]. Em problemas de classificação de imagens, como é o caso desse trabalho, essas operações geralmente são rotações, translação, ampliação, alteração do contraste, alteração do brilho, dentre outras. A ideia é que, ao aplicar essas operações nos dados, o algoritmo de classificação terá mais exemplos sendo apresentados no treinamento, e com uma variedade maior, fazendo com que ele aprenda a extrair características ainda mais relevantes sobre cada dado, para fazer a classificação.

Outra técnica de aumentação que ganhou atenção nos últimos anos é a técnica de *Auto Augmentation* [16]. Nesta, um algoritmo de aprendizado por reforço é utilizado para encontrar as transformações e parâmetros ótimos para o conjunto de dados utilizado, em oposto à simplesmente definir manualmente quais transformações as imagens irão sofrer no processo de aumento de dados. A implementação deste algoritmo é feita de tal forma que, cada política consiste em multiplicas sub-políticas, as quais contêm duas operações de transformação de imagens, como translação, rotação, expansão, dentre outras. Uma vez encontrada a política ótima para o conjunto de dados, são escolhidas e aplicadas operações de uma sub-política aleatória para cada imagem, em cada *batch* de dados do conjunto de treinamento.

D. Otimizadores

Dois otimizadores comumente utilizados na literatura são o *Stochastic gradient descent* (SGD) [17] e AdamW [18].

Ambos atualmente são escolhas comuns entre pesquisadores que buscam bater o estado da arte nos mais diversos problemas, e geralmente, quando seus parâmetros são bem definidos, possuem resultados similares. Vale notar que AdamW é uma modificação da implementação original do Adam, a qual faz o uso de *weight decay regularization* no lugar de L^2-*regularization* [19], nas equações de atualizações de peso das redes neurais.

E. Model Averaging

Model Averaging é uma técnica interessante para fortificação de resultados de modelos de aprendizado de máquina, e pode ser realizada pelo método *Stochastic Weight Averaging (SWA)* [20]. A ideia se consiste em armazenar pesos da rede neural durante o treinamento e ao fim do treinamento substituir os pesos pela média dos *samples* obtidos.

A técnica é inspirada no estudo de otimização convexa onde o objetivo, por exemplo em um cenário de duas dimensões, é sempre de buscar descer/otimizar curvas. Aplicar SWA em soluções com redes neurais que geralmente podem possuir conjuntos de pesos formando espaços com muitas dimensões e geralmente complexos (muitos mínimos locais), pelo grande número de neurônios e camadas usadas, pode causar em alguns casos o efeito da rede neural ter resultados um pouco piores no conjunto observado (treino, validação), mas que por conta da "generalização"causada pela média das redes, formada pelos conjuntos de pesos, obtidos durante o treinamento, obter resultados melhores em dados ainda não observados (teste) comparando quando não se recorra à técnica de SWA.

F. Spatial Pyramid Pooling

Spatial Pyramid Pooling é uma estratégia eficaz na transformação de um sinal convolucional espacial em um vetor de características de tamanho fixo, necessárias para a conexão com camadas densas de classificação, regressão ou outras atividades de aprendizado [21]. Dado uma sequência de inteiros que descrevem a subdivisão em malha do sinal de ativação (denominados *bins*), esta estratégia de *pooling* consiste em extrair múltiplos vetores de *features* a partir de operações usuais de *pooling* (*max-pooling* é utilizado no artigo) restritas à sub-regiões/escala determinada pelo *bin*. Esta técnica, portanto, resulta em descritores do sinal de entrada relativos à múltiplas regiões e escalas distintas a partir de um único sinal convolucional, sendo frequentemente empregada de modo a melhorar a acurácia de sistemas detectores de objetos pequenos e fora de foco [22], [23].

VI. Metodologia

Nesta seção, descrevemos nossa estratégia para o problema. Apresentamos o pré-processamento realizado nos dados, bem como cada experimento conduzido, descrevendo a exploração realizada, sua configuração e hiper-parâmetros definidos.

A. Pré-processamento

Embora a competição disponibilize dois conjuntos, um composto de imagens RGB e outro com imagens RGB-NIR (os canais usuais RGB e o canal *near-infrared*), optamos por utilizar somente as imagens RGB. Esta decisão foi tomada levando em consideração o grande número de alertas a respeito de problemas de rotulagem nas amostras RGB-NIR na competição[3].

As amostras RGB, armazenadas como imagens comprimidas no formato JPEG, foram baixadas do Kaggle e transformadas no formato TF-Records[4], que armazena as amostras (e anotações associadas) em formato binário e aleatorizado, promovendo um aumento de performance durante o treinamento e inferência sobre estes dados. Este conjunto pré-processado foi criado a fim de promover a reprodutibilidade dos resultados[5].

[3] 32453, 35915, 36787
[4] tensorflow.org/tutorials/load_data/tfrecord
[5] drive.google.com/drive/folders/1tmlk6v-t4WfdcEex25rNAelWnrcPXdAV

B. Baseline

Uma solução foi determinada como *baseline* aplicando *Transfer Learning* [13], [24] ao utilizar a rede neural *EfficientNet-B3* [25] com o modelo pré-treinado (também chamado no texto de *backbone*). Nessa implementação, é definido um bloco com uma camada de *GlobalAveragePooling2D* e uma camada densa de tamanho igual ao número de classes no problema (bloco também chamado no texto de *classification-head*) com sua entrada ligada a *EfficientNet-B3*, por sua vez com camadas congeladas (*frozen*). A *classification-head* tem seus pesos treinados primeiramente por até 80 épocas (caso não ocorra *early-stopping*). Após isso os pesos da *EfficientNet-B3* são descongelados e é realizada outra etapa de treinamento, agora atualizando pesos tanto para o modelo *backbone* quando para a *classification-head* por até mais 80 épocas, essa etapa também é chamada no texto de *fine-tuning*.

C. Data Balancing

O alto nível de desbalanceamento deste conjunto atrapalha o treinamento efetivo dos diferentes classificadores para cada rótulo, onde poucas amostras positivas para rótulos minoritárias são apresentadas (relativo ao grande número de amostras para os demais rótulos). A fim de combater essa distribuição desproporcional, diferentes estratégias de balanceamento de dados podem ser consideradas [26]. Neste trabalho, os rótulos foram subdivididos em dois grupos, dominantes e dominados. As *labels agriculture*, *clear* e *primary* foram alocadas no grupo das dominantes, enquanto as demais foram assinaladas como dominadas. O conjunto foi então segmentado por rótulo. Para cada rótulo no conjunto de dominadas, todas as amostras associadas ao rótulo foram coletadas e subdivididas em sub-conjuntos correspondentes. Para cada rótulo do conjunto de dominantes, as amostras com no máximo uma *label* ocorrente (a dominante) foram utilizadas. Um novo conjunto foi então formado a partir da amostragem uniforme a partir de todos os sub-conjuntos formados.

A amostragem foi feita com o auxílio da função `tf.data.experimental.sample_from_datasets`, que simplifica a operação de "custura" entre múltiplos conjuntos.

D. Data Augmentation

Foram testadas duas formas de fazer o aumento artificial dos dados, na primeira tentativa, foram escolhidas manualmente 6 operações sendo elas: espelhar horizontalmente a imagem, espelhar verticalmente a imagem, alterar a matiz, alterar o brilho, alterar o contraste e alterar a saturação da imagem.

Na segunda tentativa, foi utilizado a técnica *Auto Augmentation*. Como *Auto Augmentation* exige um treinamento prévio com aprendizagem por reforço, o que pode levar um tempo considerável, foi utilizado duas políticas pré-treinadas, uma na base dados ImageNet e outra no CIFAR-10. A biblioteca *DeepVoltaire*[6] foi reutilizada para simplificar a aplicação desta técnica.

[6] github.com/DeepVoltaire/AutoAugment

E. Otimizadores

Transfer learning foi utilizado para treinar nossos modelos em dois estágios. No primeiro, os pesos das camadas do modelo *backbone* eram congelados (o vetor gradiente da função de perda com respeito aos pesos era anulado) e a camada de classificação era treinada. Em um segundo passo, 60% das camadas superiores do *backbone* eram descongeladas e treinadas novamente (juntamente com a camada de classificação), utilizando um *learning rate* reduzido.

Dois experimentos foram realizados. No primeiro, utilizamos o otimizador SGD com *learning rate* $0,1$ e Nesterov momentum igual à $0,9$ (valores razoáveis encontrados na literatura) para o treinamento de nosso modelo no primeiro estágio, e *learning rate* igual à $0,01$ no segundo estágio. Em um segundo experimento, utilizamos o otimizador AdamW com *learning rate* $0,001$ e *weight decay* igual à $0,0001$ no primeiro estágio, e *learning rate* igual à $0,0001$ no segundo estágio.

F. Focal Loss

Utilizamos a função de perda denominada *focal loss* [27] para abordar o problema de desbalanceamento de classes durante o treinamento, visando diminuir o viés da classe dominante. Esta técnica consiste em usar um parâmetro de modulação para dar maior foco no aprendizado de amostras difíceis, e é definida da da seguinte forma:

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{2}$$

G. Spatial Pyramid Pooling

Neste experimento, a rede *EfficientNet-B3* é utilizada como *backbone*. O sinal proveniente das camadas denominadas *block5a_expand_activation*, *block6a_expand_activation* e *top_activation* são extraídos e *Spatial Pyramid Pooling* [21] com bins 1, 3 e 5 é utilizado para extrair um sinal de *features* de formato $(B, 35, K)$ de cada um dos sinais. *Max Average Pooling 1D* é então utilizado para transformar este sinal em vetores de *features*, que são concatenadas em um único vetor descritor da amostra.

H. Model Averaging

Utilizamos SWA em experimentos de modo a fortificar nossos resultados sobre o conjunto de teste, e na esperança de proporcionar maior generalização para predição de dados não vistos. Nas duas etapas, (i) treinamento da *classification-head* com o modelo *backbone* congelado e (ii) *fine-tuning* com *backbone* descongelado, os pesos foram salvos a cada 10 épocas a partir da primeira época, e ao final do processo de treinamento os pesos da rede foram atualizados com a média resultante daqueles armazenados em cada respectiva posição.

VII. Resultados e Discussão

Nesta seção são apresentados os resultados dos experimentos conduzidos, assim como uma discussão sobre estes. Além disso, em anexo a este trabalho, se encontra uma planilha com as métricas detalhadas de cada experimento.

Tabela I
F_2-SCORE POR RÓTULO E ESTRATÉGIA SOBRE O CONJUNTO DE VALIDAÇÃO E TESTE.

Label	EB3 (*baseline*)	EB3-B	EB3-SPP	EB3-AdamW	EB3-AA-IN-8	EB3-FL	EB3-SWA
agriculture	**84.69%**	73.56%	83.21%	83.86%	83.39%	61.55%	84.12%
artisinal_mine	80.87%	**90.17%**	83.49%	77.15%	74.60%	63.79%	80.36%
bare_ground	12.52%	15.74%	**24.16%**	11.09%	7.40%	0.00%	12.52%
blooming	0.00%	**59.03%**	0.00%	0.00%	0.00%	0.00%	0.00%
blow_down	0.00%	**35.26%**	0.00%	0.00%	0.00%	0.00%	0.00%
clear	96.87%	96.50%	96.63%	**97.20%**	97.14%	95.62%	96.84%
cloudy	85.23%	**88.91%**	86.41%	83.31%	80.16%	78.91%	84.60%
conventional_mine	0.00%	**76.09%**	21.05%	0.00%	0.00%	0.00%	0.00%
cultivation	48.45%	52.67%	51.41%	41.75%	**57.93%**	12.64%	46.72%
habitation	63.12%	63.86%	60.91%	60.06%	**66.69%**	20.50%	63.35%
haze	66.14%	**66.16%**	65.27%	62.90%	65.13%	50.29%	**66.48%**
partly_cloudy	**91.62%**	89.78%	90.70%	89.52%	90.20%	83.97%	**91.80%**
primary	**98.56%**	98.48%	98.23%	98.54%	**98.82%**	97.70%	98.69%
road	82.99%	81.35%	83.74%	82.84%	**86.02%**	63.72%	82.79%
selective_logging	3.70%	**38.72%**	18.39%	0.00%	0.00%	0.00%	3.69%
slash_burn	0.00%	**17.72%**	0.00%	0.00%	0.00%	0.00%	0.00%
water	**76.44%**	73.79%	76.12%	71.90%	71.60%	55.02%	**77.13%**
Valid média simples	52.42%	**65.75%**	55.28%	50.60%	51.71%	40.22%	52.30%
Valid média ponderada	88.14%	80.39%	87.98%	87.20%	**88.26%**	78.55%	88.09%
Test public	90.06%	88.05%	90.06%	89.41%	89.98%	81.09%	**90.14%**
Test private	89.78%	87.80%	89.83%	89.22%	89.65%	80.61%	**89.90%**

A Tabela I apresenta os resultados da métrica F_2-score de cada estratégia sobre o conjunto Planet: Understanding the Amazon from Space. As 20 primeiras linhas exibem os resultados por *label* sobre o conjunto de validação, enquanto a vigésima primeira e vigésima segunda linhas exibem suas médias simples e ponderadas pela frequência das classes, respectivamente. Por fim, as duas últimas linhas apresentam os resultados sobre o conjunto de teste público e privado no Kaggle, respectivamente.

1) Baseline: a pontuação do modelo *baseline* (EB3) é exibida na primeira coluna da Tabela I. Observamos que este treinamento resultou nos melhores resultados para as labels *agriculture*, *partly cloudy*, *primary* e *water*, porém vários rótulos obtiveram F_2 de 0%, indicando que a rede está ignorando classes sub-representadas e se focando na identificação das classes majoritárias.

2) Data Balancing: resultados para o treinamento após o balanceamento de dados (EBC3-B) estão expressos na segunda coluna da Tabela I. O balanceamento prévio dos dados promoveu uma melhora significativa na pontuação das classes sub-representadas, que foi traduzido em uma melhora de $13,33$ pontos percentuais em relação ao baseline (EB3). Por outro lado, a pontuação utilizada internamente no Kaggle é a F_2-*score* ponderada. Esta estratégia, portanto, não alcançou uma melhora em relação ao *baseline*.

3) Spatial Pyramid Pooling: é possível observar que a estratégia de *Spatial Pyramid Pooling* (EB3-SSP) um aumento na métrica F_2-*score* para todos os rótulos associados à pequenos elementos (relativos ao tamanho total da imagem), como *artisinal mine*, *conventional mine*, *cultivation*, *road*, *selective logging*, etc. Em média, $F_2 - macro$ é $2,86$ pontos percentuais acima do *baseline*. Por outro lado, observamos uma leve redução de $0,16$ pontos percentuais em $F_2 - weighted$, comparado ao *baseline*.

4) Otimizadores: foi testado o uso de SGD com nesterov momentum (EB3 *baseline*) e AdamW (EB3-AdamW) como otimizadores, e em nossos experimentos, SGD com nesterov momentum apresentou os melhores resultados, com *score* de

$90,06\%$, enquanto AdamW obteve no melhor dos casos um *score* de $89,41\%$. Os *scores* encontrados foram bem próximos, o que pode ser um indício de que caso seja encontrado o conjunto de hiperparâmetros ótimos para AdamW neste problema, é possível que ele tenha resultado similar ou que seja até melhor que SGD com nesterov momentum.

5) Data Augmentation: em nosso *baseline*, um experimento onde a aumentação dados é realizada utilizando operações de transformação de imagem definidas manualmente, foi obtido um *score* de $90,06\%$, sendo esse o melhor resultado dentre as duas formas de aumento de dados testadas. Usando o *Auto Augmentation* (EB3-AA-IN-8), com uma política pré-treinada na base ImageNet, foi realizado um experimento gerando 4 novas imagens e 8 novas imagens, resultando em um *score* de $89,82\%$ e $89,98\%$ respectivamente. Com uma política pré-treinada na base CIFAR-10, gerando 4 novas imagens, o *score* foi de $89,45\%$. Ao contrário do que se esperava, a escolha manual de operações de transformações de imagem foi o método que apresentou melhor resultado. Isso provavelmente se dá, pois, o *Auto Augmentation* foi treinado para lidar com imagens do conjunto de dados ImageNet e CIFAR-10, os quais possuem imagens consideravelmente diferentes das utilizadas neste trabalho, e, as transformações escolhidas foram manualmente definidas ao se basear em conhecimento sobre o domínio do problema proposto neste projeto.

Nos experimentos onde foram gerados 4 e 8 novas imagens a partir de transformações na imagem original usando a política do Auto Augmentation, o melhor resultado foi obtido ao gerar 8 novas imagens. Esse é um resultado esperado, pois dessa forma o algoritmo tem acesso a mais exemplos, e com uma variedade maior, durante seu treinamento, porém, como a diferença de *score* foi muito pequena e o tempo de treinamento aumenta consideravelmente ao adicionar mais imagens no conjunto, seu uso não é uma alternativa viável durante o processo de desenvolvimento do algoritmo.

6) Focal Loss: levando em consideração o desbalanceamento das amostras foram feitos dois experimentos com a *focal loss* como função de otimização (EB3-FL), usando dois valores diferentes para o termo modulador: $\gamma = 1$ e $\gamma = 2$. Em ambos os casos o resultado não foi melhorado em consequência do baixo desempenho em prever *true positives* e *false negatives*, tendo como resultado os *scores* de $80,61\%$ e $82,35\%$ para o teste privado e $81,09\%$ e $82,86\%$ para o teste público, para o parâmetro $\gamma = 1$ e $\gamma = 2$, respectivamente.

7) Model Averaging: partindo do modelo *baseline* foram feitos experimentos aplicando SWA para *sampling* de pesos (i) durante apenas a etapa de *fine-tuning* onde o modelo *backbone* teve seus pesos descongelados, e (ii) desde a etapa de treinamento do *classification-head* até o fim da etapa de *fine-tuning*. A primeira maneira, aplicando-se apenas no *fine-tuning*, não obteve resultados melhores que o do modelo *baseline*, já a segunda teve resultados próximos no conjunto de validação e superou os demais métodos explorados no conjunto de teste privado e público.

O método por SWA aparece na Tabela I com rótulo "EB3-SWA", observa-se que os resultados não são destacantes em relação ao modelo *baseline* quando se consideram as métricas para cada classe no conjunto de validação. No entanto, o modelo apresenta nos conjuntos de teste os melhores resultados de $90,14\%$ para o conjunto de teste público, e $89,90\%$ para o conjunto de teste privado.

Por fim, foi observado a ocorrência do efeito possível e característico do SWA de produzir resultados próximos ou piores que o seu não uso no conjunto de validação e obter melhorias no conjunto de dados de teste. Comportamento que pode ser explicado pela esperança do método de proporcionar generalização quando considerados dados desconhecidos, já que a rede resultante final foi criada pela média dos pesos durante o treinamento, estas que formavam as melhores redes em sua respectiva época de *sampling*.

VIII. Conclusões

Neste trabalho, investigamos a aplicação de técnicas de aprendizado profundo à solução do desafio "Planet: Understanding the Amazon from Space". Investigamos o efeito de diversas técnicas relacionadas sobre um baseline estabelecido (*EfficientNet-B3*), como o balanceamento de dados, *Spatial Pyramid Pooling*, a utilização de múltiplos otimizadores, técnicas de aumentação de dados e, finalmente, a combinação de múltiplos modelos para a fortificação das respostas.

Quando avaliadas sob as mesmas condições impostas durante a competição, observou-se que a combinação de múltiplos modelos utilizando a estratégia SWA obteve a melhor pontuação sobre os conjuntos de teste privado e público, $90,14\%$ e $89,90\%$, respectivamente.

Embora estes resultados sejam encorajadores, eles não se aproximaram das primeiras posições no quadro de liderança, estando 3,42 pontos percentuais abaixo do ganhador da competição. O trabalho estudado aqui pode ser futuramente estendido (e resultados possivelmente melhorados) com a utilização da informação *near-infrared*, com um estudo detalhado do ruído sobre o conjunto de rótulos no conjunto de treinamento e a utilização de técnicas de remoção de ruído [28] e neblina [29].

Referências

[1] L. C. Barbosa, *The Brazilian Amazon rainforest: Global ecopolitics, development, and democracy*. University Press of America, 2000. 1

[2] J. F. do Vale Júnior, M. I. L. de Souza, P. P. R. Nascimento, and D. L. de Souza Cruz, "Solos da amazônia: etnopedologia e desenvolvimento sustentável," *Revista Agro@ mbiente On-line*, vol. 5, no. 2, pp. 158–165, 2011. 1

[3] "Planet: Understanding the amazon from space," 2017. [Online]. Available: https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/overview 1

[4] "Monitoramento da floresta amazônica brasileira por satélite," São José dos Campos, 2011. [Online]. Available: http://www.obt.inpe.br/prodes/ 1

[5] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *European conference on information retrieval (ECIR)*. Springer, 2005, pp. 345–359. 2

[6] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "A comparison of multi-label feature selection methods using the problem transformation approach," *Electronic Notes in Theoretical Computer Science*, vol. 292, pp. 135–151, 2013. 2

[7] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, 2017. 3

[8] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence (PAI)*, vol. 9, no. 2, pp. 85–112, 2020. 3

[9] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, pp. 1–1, 2021. 3

[10] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4724–4732. 3

[11] J. Pons, O. Slizovskaia, R. Gong, E. Gómez, and X. Serra, "Timbre analysis of music audio signals with convolutional neural networks," in *25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 2744–2748. 3

[12] L. Chan, M. S. Hosseini, and K. N. Plataniotis, "A comprehensive analysis of weakly-supervised semantic segmentation in different image domains," *International Journal of Computer Vision (IJCV)*, vol. 129, no. 2, pp. 361–384, 2021. 3

[13] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *ICML workshop on unsupervised and transfer learning*. JMLR Workshop and Conference Proceedings, 2012, pp. 17–36. 3, 5

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015. 4

[15] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019. 4

[16] E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 113–123. 4

[17] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *30th International Conference on Machine Learning (ICML)*, ser. Proceedings of Machine Learning Research, S. Dasgupta and D. McAllester, Eds., vol. 28, no. 3. Atlanta, Georgia, USA: PMLR, 17–19 Jun 2013, pp. 1139–1147. [Online]. Available: http://proceedings.mlr.press/v28/sutskever13.html 4

[18] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *ICLR*, 2019. 4

[19] T. Van Laarhoven, "L2 regularization versus batch and weight normalization," *arXiv preprint arXiv:1706.05350*, 2017. 4

[20] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A. G. Wilson, "Averaging weights leads to wider optima and better generalization," *arXiv preprint arXiv:1803.05407*, 2018. 4

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015. 4, 5

[22] K.-J. Kim, P.-K. Kim, Y.-S. Chung, and D.-H. Choi, "Performance enhancement of yolov3 by adding prediction layers with spatial pyramid pooling for vehicle detection," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2018, pp. 1–6. 4

[23] Z. Huang, J. Wang, X. Fu, T. Yu, Y. Guo, and R. Wang, "Dc-spp-yolo: Dense connection and spatial pyramid pooling based yolo for object detection," *Information Sciences*, vol. 522, pp. 241–258, 2020. 4

[24] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016. 5

[25] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning (ICML)*. PMLR, 2019, pp. 6105–6114. 5

[26] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004. 5

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988. 5

[28] Y. Zhang and R. K. Mishra, "A review and comparison of commercially available pan-sharpening techniques for high resolution satellite image fusion," in *IEEE International geoscience and remote sensing symposium*. IEEE, 2012, pp. 182–185. 7

[29] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010. 7