

Prova de conceito de um modelo classificador de áreas irregulares na Floresta Amazônica baseado em Transformers Visuais

Victor Moraes
15 de dezembro de 2022

UFMG

Estrutura da Apresentação

Introdução

Motivação

Revisão Bibliográfica

Metodologia

Análise exploratória

Resultados

Conclusão

Referências

Introdução

Motivação





Figura 1:

Garimpo ilegal na Terra Indígena Munduruku, município de Jacareacanga. Foto: Marizilda Cruppe/Amazônia Real

Revisão Bibliográfica

- Domínio do problema: Sensoriamento remoto
- Aprendizado de máquina
- Redes convolucionais
- Transformers
- Trabalhos anteriores

Revisão bibliográfica - Sensoriamento Remoto

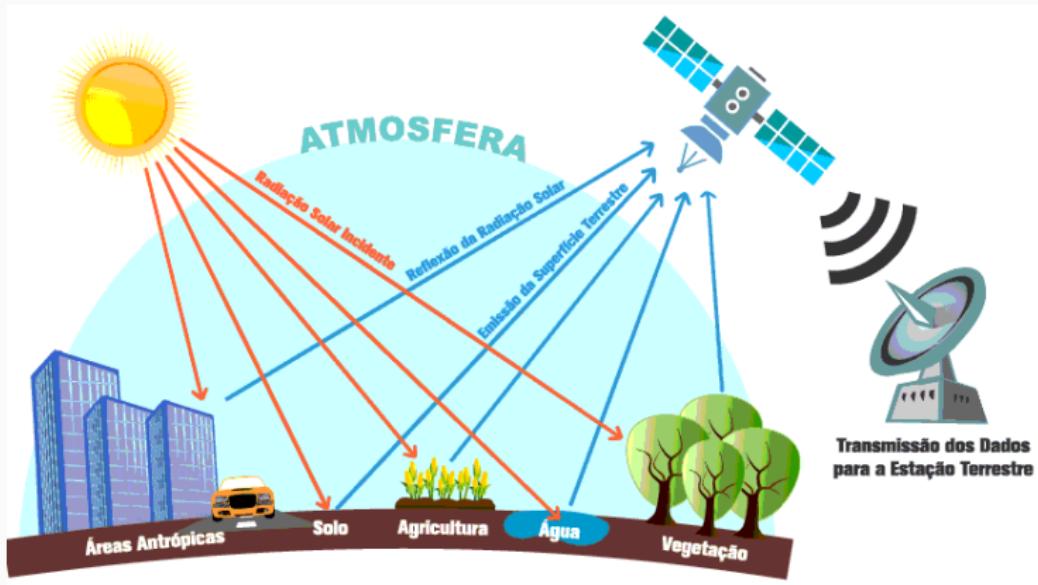


Figura 2: Sensoriamento remoto Foto:(INSTRUTORGIS, 2022)

“Campo de estudos que visa a dar computadores a habilidade de aprender sem serem explicitamente programados para determinada tarefa.” (SAMUEL, 1959)

“Um algoritmo dito conseguir uma experiência E com respeito a determinada classe de tarefas T e com medidas de desempenho P, se seu desempenho nas tarefas em T, medidas por P, melhoram a partir da experiência E.” (MITCHELL, 1997)

Aprendizado de máquina - Exemplo

Exemplo:

- Tarefa T: Problema de classificação
- Experiência E: Amostras dentre k classes a classificar
- Desempenho P: acurácia do algorítimo

Producir função:

$$f : \Re^n \rightarrow \{1, \dots, k\}. \text{ Quando } y = f(x)$$

Métricas de desempenho

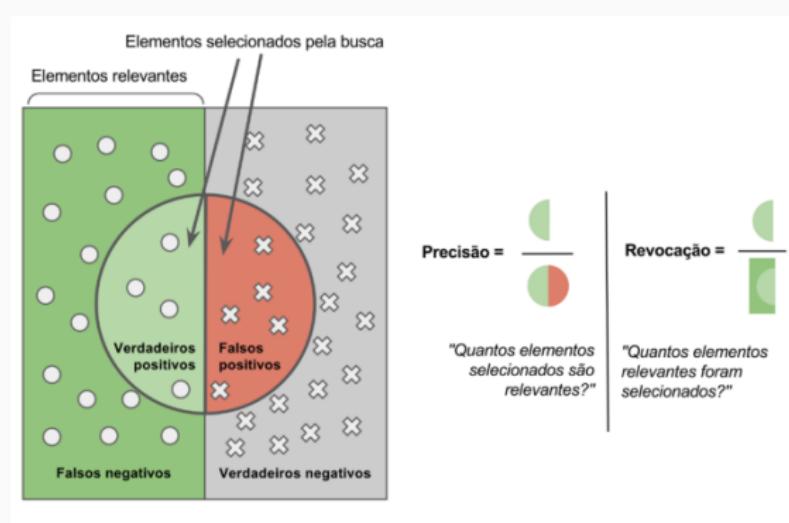


Figura 3: Ilustração das métricas de precisão e revocação. Fonte: WikiCommons

Métricas de desempenho

		Detectada	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

$$P = \frac{VP}{VP + FN} \quad (1)$$

$$R = \frac{VP}{VP + FP} \quad (2)$$

$$AC = \frac{VP + VN}{VP + VN + FN + FP} \quad (3)$$

Figura 4: Matriz Confusão, Equações de Precisão, Revocação e Acurácia

Métricas de desempenho

$$F_1 = \frac{P \times R}{P + R} \quad (4)$$

$$F_\beta = (1 + \beta^2) \times \frac{P \times R}{\beta^2 P + R} \quad (5)$$

$$F_2 = 5 \times \frac{P \times R}{5P + R} \quad (6)$$

Figura 5: Equações de F_β e F_2

Curva PR

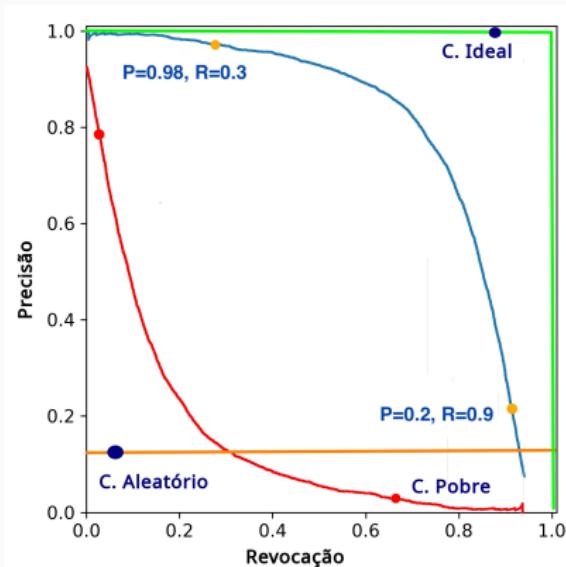
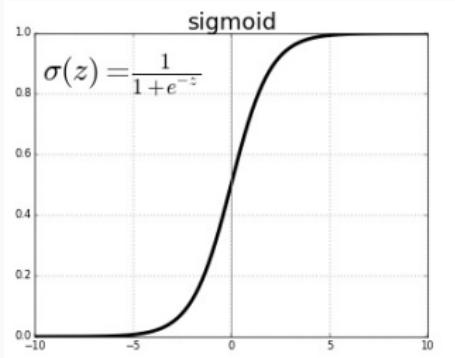


Figura 6: Curva Precisão-Recisão

Regressão logística



$$\sigma(x) = \frac{1}{1 + \exp(-z(x))} \quad (7)$$

Figura 7: Função de ativação *Sigmoid*.

$$w \leftarrow w - \eta \nabla L(w) \quad (8)$$

$$Q(w) = \frac{1}{n} \sum_i Q_i(w) \quad \Rightarrow \quad \nabla Q(w) = \frac{1}{n} \sum_i \nabla Q_i(w) \quad (9)$$

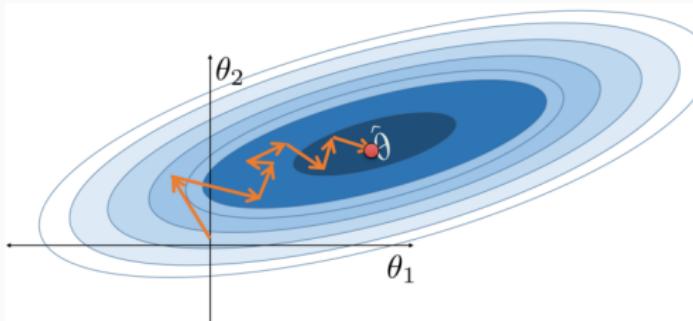


Figura 8: Ilustração do algoritmo gradiente descendente estocástico.

Função de perda

$$ECB(y_c) = - \sum_{c=0}^M y_c \times \log(\hat{y}_c) \quad (10)$$

$$Focal(y_c) = - \sum_{c=0}^M (1 - w_i \times y_c)^\gamma \times \log(\hat{y}_c) \quad (11)$$

Redes Neurais Artificiais Profundas

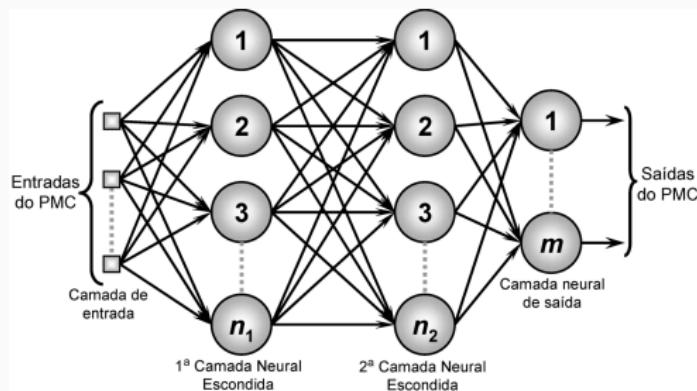


Figura 9: Exemplo de uma rede de perceptrons de multiplas camadas - MLP.

Redes convolucionais

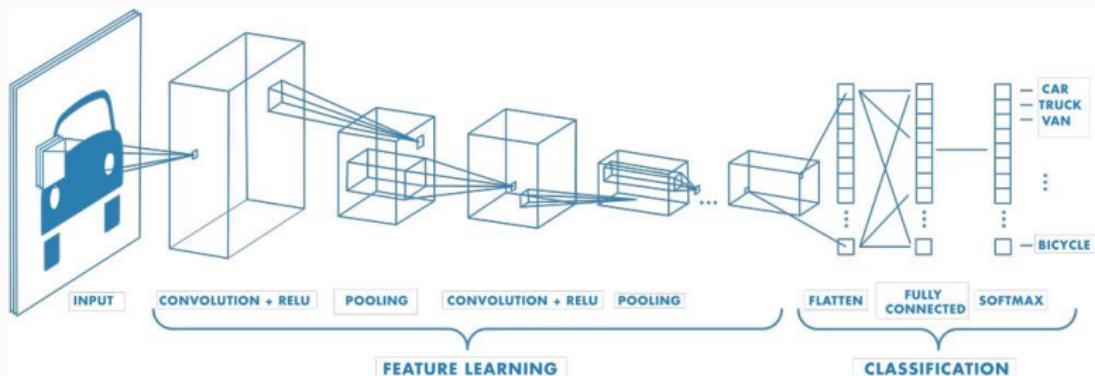


Figura 10: Arquitetura de uma rede convolucional.

Redes convolucionais

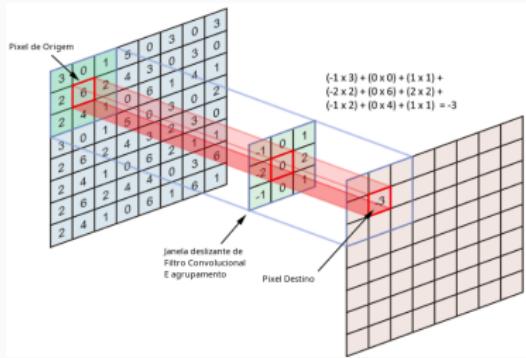


Figura 11: Operação de janela deslizante aplicando operações de convolução e pooling.

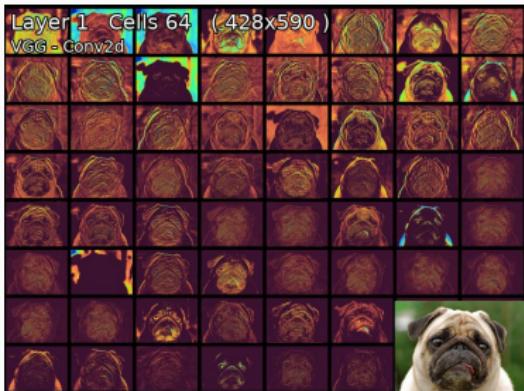


Figura 12: Biblioteca Pytorch de visualização de filtros extratores.

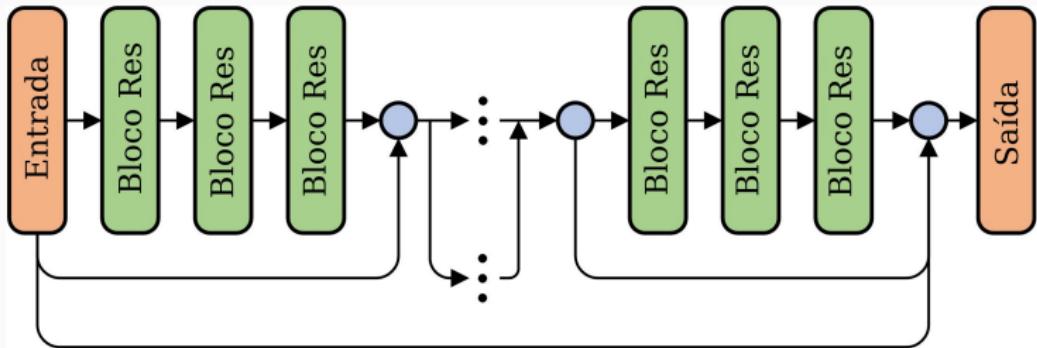


Figura 13: Arquitetura redes ResNet.

Transformer Visual

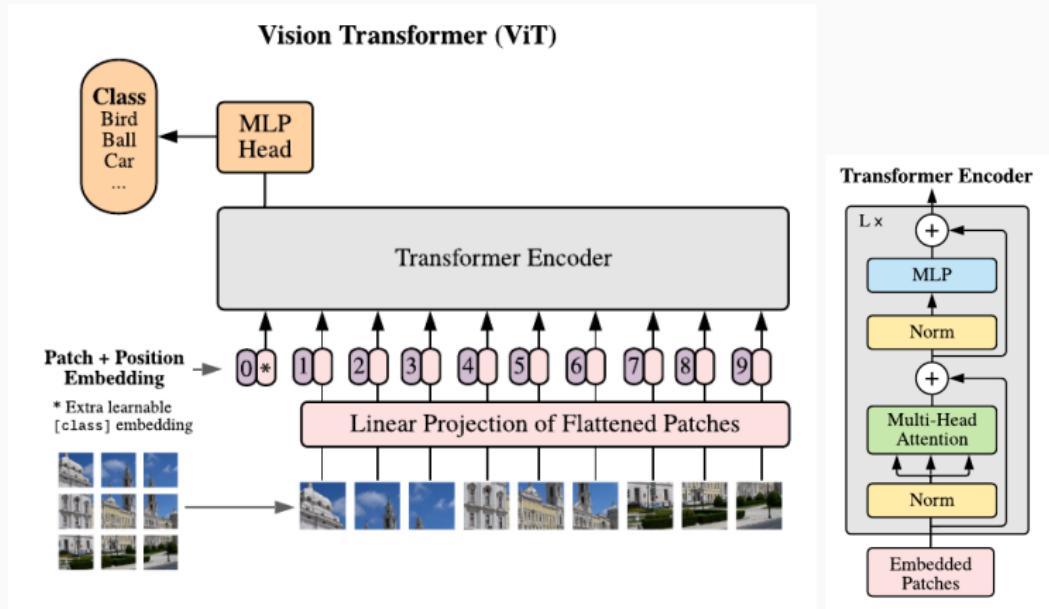
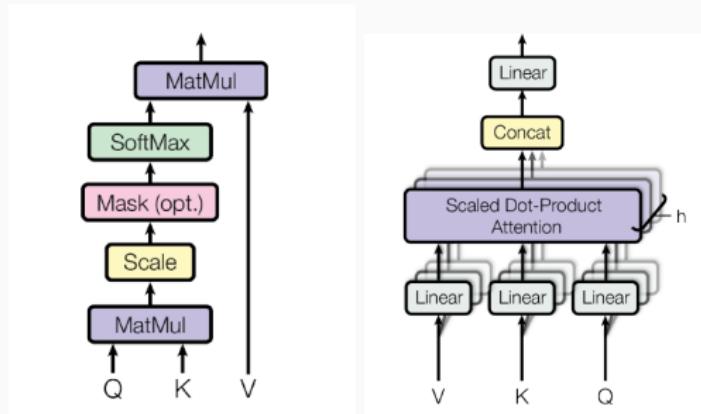


Figura 14: (DOSOVITSKIY et al., 2020)

Auto-Atenção



$$Y = \text{softmax}(QK^T)V \quad (12)$$

Figura 15: (DOSOVITSKIY et al., 2020)

Transformer Swin

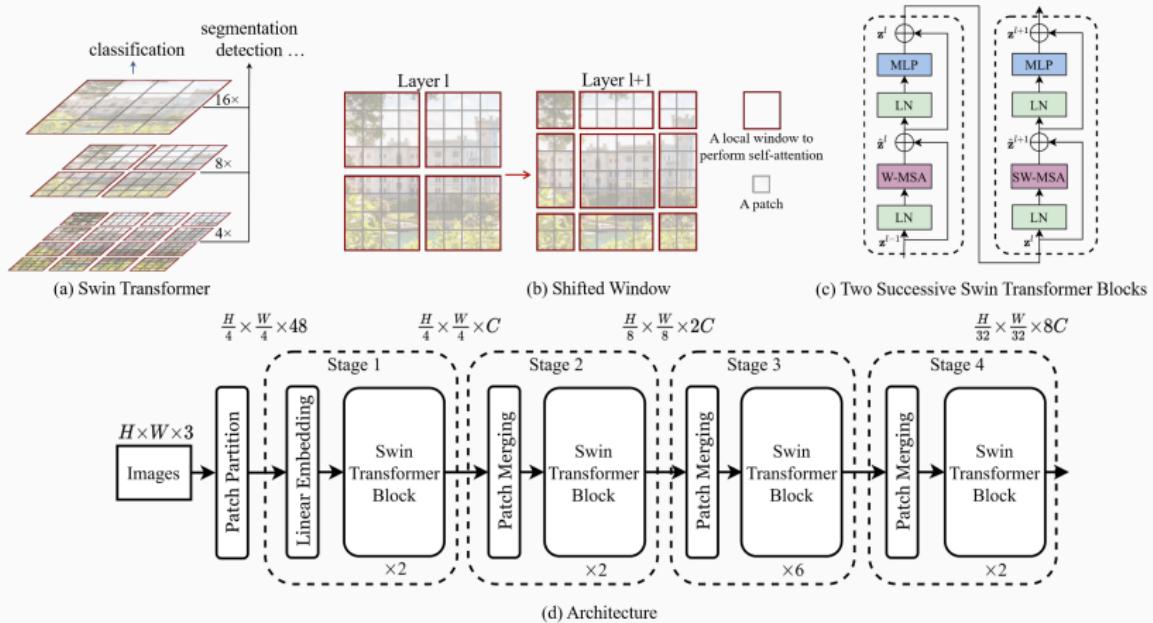


Figura 16: Arquitetura Swin (*Shifted-Window* - Janela deslocada) (LIU et al., 2022)

Trabalhos anteriores

- Detecção de mudanças para detecção de minas de ouro (CAMALAN et al., 2022)
- Classificação de minas e represas (BALANIUK; ISUPOVA; REECE, 2020)
- Dataset MillionAID e transformers visuais (WANG et al., 2022)
- transformers visuais para detecção de desmatamento (KASELIMI et al., 2022)

Metodología

Método



Figura 17: Amostras de classes do dataset Amazônia do Espaço Fonte:(PLANETCO, 2016)

Premissas e proposta de solução

- Explorar viés indutivo dos transformers visuais
- Comparar desempenho de um modelo Swin e ResNet

Ambiente e ferramentas

- Jupyter e ambiente cloud Google Collab
- Pytorch, SciKitLearn e NumPy

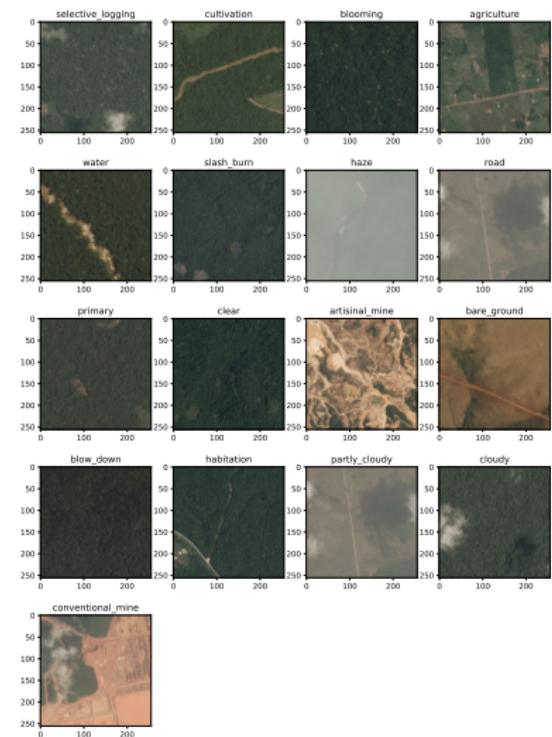
Experimentos

1. Análise exploratória dos dados.
2. Pre-processamento do conjunto de dados e de carregadores de amostras.
3. Elaborar um modelo base.
4. Utilizar pesos dos trabalhos disponibilizados para fine-tune.
5. Elaborar modelo proposto baseado em transformers visuais.
6. Fixar os hiperparâmetros e metodologia de treino para modelos base e proposto e treiná-los novamente a fim de realizar comparações.
7. Análise dos resultados de ambos comparando desempenho em métricas relevantes ao problema.

Análise exploratória

Classe	Rótulo	Amostras	Proporção (%)
Mina Convencional	conventional mine	100	0,247
Roça de Ventos	blow down	101	0,250
Queimada	slash burn	209	0,516
Florescimento	blooming	332	0,820
Garimpo	artisinal mine	339	0,837
Desmatamento Seletivo	selective logging	340	0,840
Área Descoberta	bare ground	862	2,129
Nublado	cloudy	2089	5,161
Névoa	haze	2697	6,663
Habitação	habitation	3660	9,042
Cultivação	cultivation	4547	11,233
Parcialmente Nublado	partly cloudy	7261	17,938
Águas	water	7411	18,308
Estrada	road	8071	19,939
Agricultura	agriculture	12315	30,423
Clima Limpo	clear	28431	70,236
Vegetação Primária	primary	37513	92,673

Amostragem de cada classe



Matriz de co-ocorrência

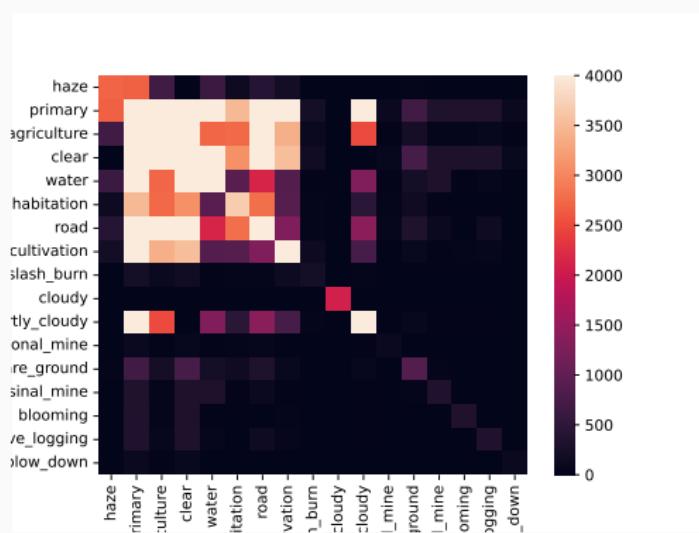


Figura 19: Matriz de co-ocorrência. Fonte: Autor

Agrupamento de amostras



Figura 20: Agrupamento de amostras via técnica TSNE. Fonte: Autor

Pré-processamento, Transformação e carregamento

1. Carregamento dos canais RGB e descartando o de infravermelho-próximo de cada imagem.
2. Redimensionamento usando interpolação linear de $256 \times 256\text{px}$ para $224 \times 224\text{px}$.
Isto se deve ao fato dos modelos já terem sido pré-treinados e configurados para essa dimensão de entrada.
3. Conversão da imagem para estrutura de dados numérica de Tensor
4. Aplicar espelho vertical ou horizontal, cada um com probabilidade de 25
5. Normalizar cada canal de cor RGB usando normalização Gaussiana com médias e desvio padrão do dataset ImageNet.

Definição do modelo

Nome	Pré-Treino	Resolution	Acurácia@1	Acurácia@5	Parâmetros	FLOPs
Swin-T	ImageNet-1K	224x224	81.474	95.5	28.3M	4.5G
Resnet50	ImageNet-1K	224x224	80.858	92.9	25.5M	4.1G

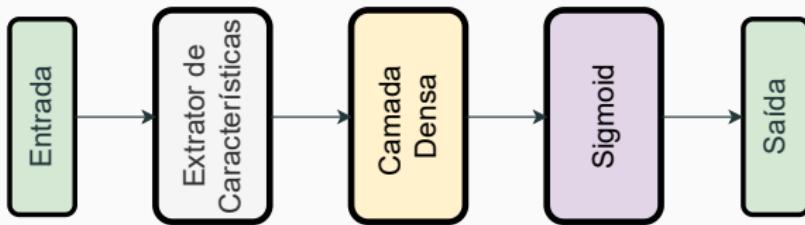


Figura 21: Definição do modelo.

Seleção de modelos

1. Pesos pré-treinados
2. Aumentar capacidade da rede
3. Adicionar regularização
4. Amostrador
5. Aumento de dados aleatória
6. Função de perda
7. Otimizador
8. Taxas de aprendizado
9. Transferência de aprendizado vs Fine Tune
10. Experimentar os mesmos passos para o modelo Swin-T

Treino e validação - Modelo base

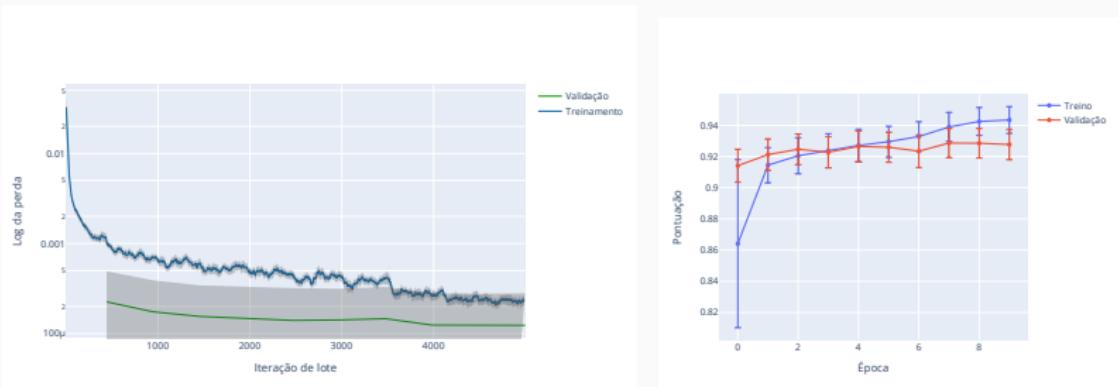


Figura 22: Métrica de perda e F2 em treino e validação por época para a rede ResNet50.

Treino e validação - Modelo proposto

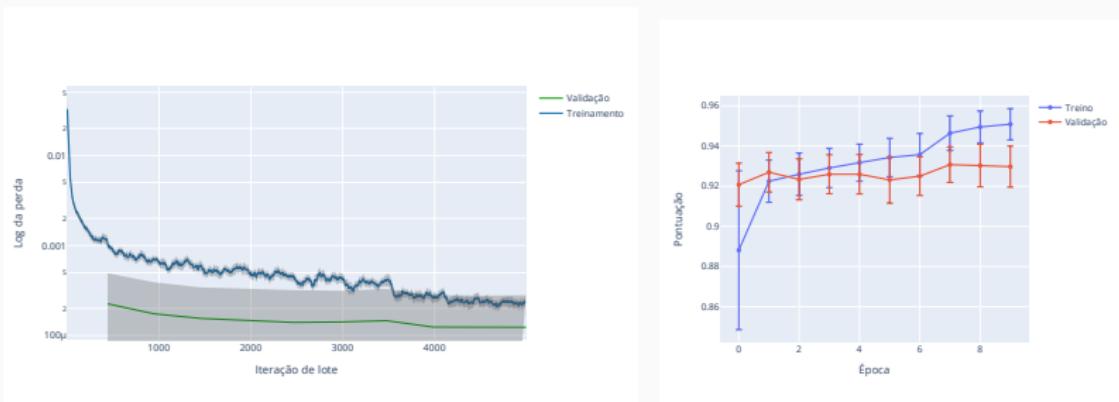


Figura 23: Métrica de perda e F2 em treino e validação por época para a rede Swin-T.

Resultados Modelo base - Matriz Confusão



Figura 24: Matriz Confusão para o modelo base Resnet-50. Fonte: Autor

Resultados

Resultados do Modelo base

	Rótulo	F2	Limiar	PR AUC	AUC Ingênuo
15	slash burn	0,030	0,230	0,145	0,005
3	blooming	0,140	0,130	0,096	0,008
4	blow down	0,268	0,190	0,245	0,003
2	bare ground	0,373	0,210	0,316	0,024
14	selective logging	0,422	0,090	0,401	0,010
7	conventional mine	0,579	0,170	0,536	0,002
8	cultivation	0,674	0,130	0,650	0,114
9	habitation	0,769	0,130	0,802	0,093
10	haze	0,774	0,210	0,784	0,063
16	water	0,836	0,210	0,892	0,181
1	artisinal mine	0,840	0,190	0,880	0,008
13	road	0,864	0,210	0,916	0,200
0	agriculture	0,890	0,250	0,929	0,306
6	cloudy	0,910	0,230	0,946	0,051
11	partly cloudy	0,938	0,210	0,972	0,173
5	clear	0,978	0,210	0,996	0,713
12	primary	0,992	0,190	0,999	0,928
17	global	0,928	0,188	0,677	0,170

Classes Raras

Classe	Rótulo	Amostras	Proporção (%)
Mina Convencional	conventional mine	100	0,247
Roça de Ventos	blow down	101	0,250
Queimada	slash burn	209	0,516
Florescimento	blooming	332	0,820
Garimpo	artisanal mine	339	0,837
Desmatamento Seletivo	selective logging	340	0,840
Área Descoberta	bare ground	862	2,129
Todo conjunto de dados	global	40479	100.0

Curva PR

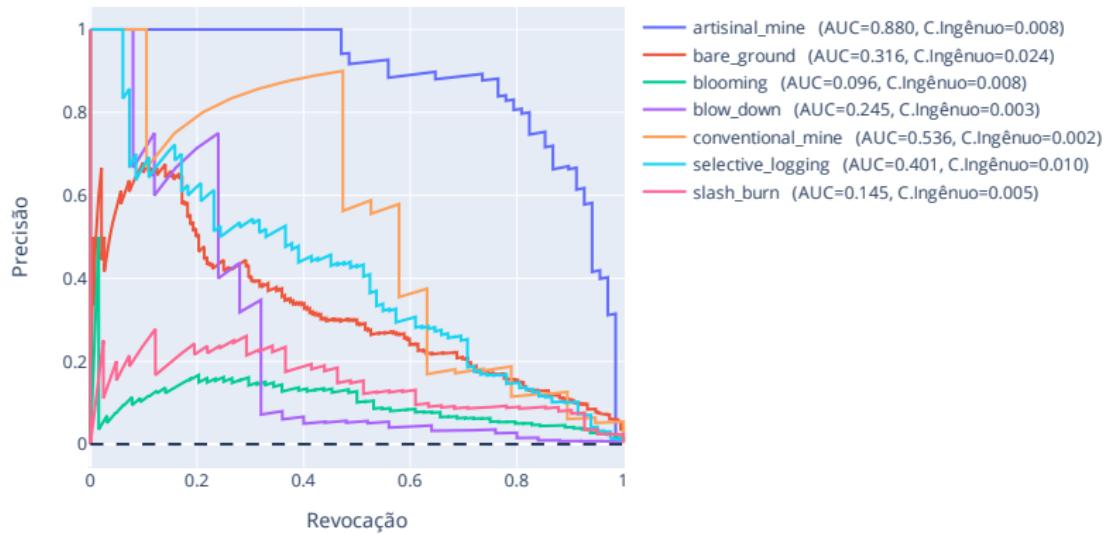


Figura 25: Curva PR para o modelo base. Fonte: Autor

Resultados modelo proposto

Rótulo	F2	PR AUC	PR-AUC	Class.Ingênuo
blooming	0,230	0,129		0,008
slash burn	0,338	0,342		0,005
blow down	0,310	0,279		0,003
bare ground	0,447	0,344		0,024
selective logging	0,475	0,422		0,010
conventional mine	0,538	0,625		0,002
artisinal mine	0,872	0,880		0,008
global	0,930	0,704		0,170

Resultados e discussão

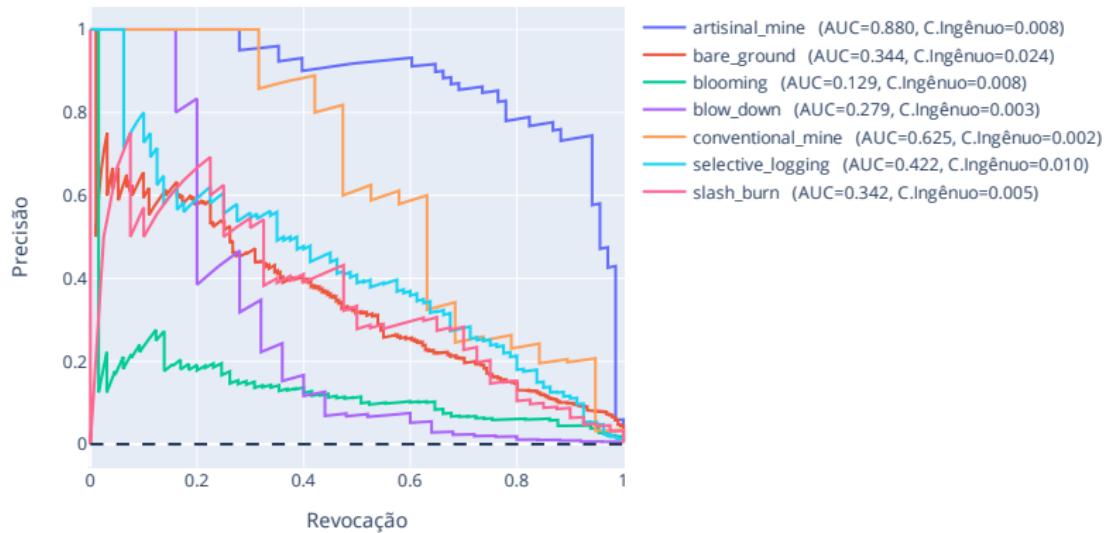


Figura 26: Curva Precisão-Revocação para o modelo proposto. Fonte: Autor

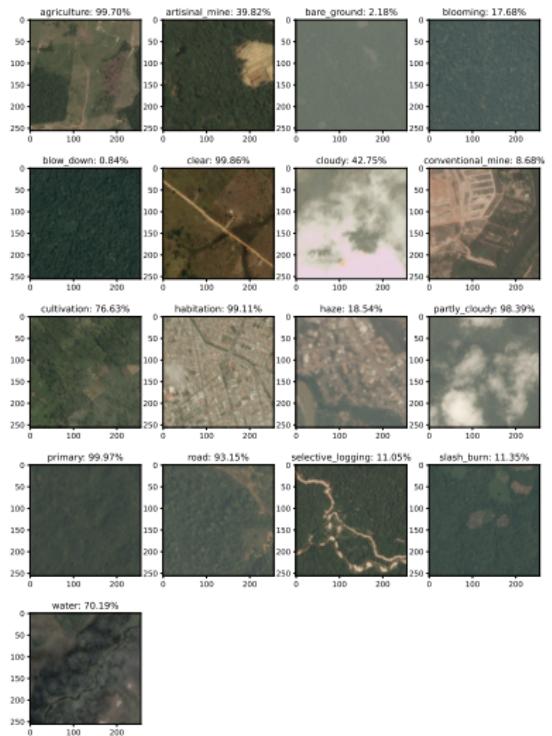
Comparação de resultados

Rótulo	F2-Resnet	F2-SwinT
slash burn	0,030	0,338
blooming	0,140	0,230
blow down	0,268	0,310
bare ground	0,373	0,447
selective logging	0,422	0,475
conventional mine	0,579	0,538
artisinal mine	0,840	0,872
global	0,928	0,930

Resultados e discussão

Rótulo	PR-AUC-ResNet	PR-AUC-SwinT	PR-AUC Class.Ingênuo
slash burn	0,145	0,342	0,005
blooming	0,096	0,129	0,008
blow down	0,245	0,279	0,003
bare ground	0,316	0,344	0,024
selective logging	0,401	0,422	0,010
conventional mine	0,536	0,625	0,002
artisinal mine	0,880	0,880	0,008
global	0,677	0,704	0,170

Resultados e discussão



Conclusão

Conclusão

- Sensoriamento remoto
- Aprendizado profundo
- Metodologia para desenvolvimento de novos modelos
- Redes CNN vs Transformers
- Prova de conceito

Trabalhos futuros

- Visualização e explicabilidade do o extrator de características
- Outros datasets para validação das conclusões
- Embutir em uma aplicação
- Perda Hamming, adequada para classificações multi-rótulos.
- Geração de dados sintéticos para classes raras utilizando ruido gaussiano.
- Canal de infravermelho próximo

Referências

 BALANIUK, R.; ISUPOVA, O.; REECE, S. Mining and tailings dam detection in satellite imagery using deep learning. *Sensors*, v. 20, n. 23, 2020. ISSN 1424-8220.

Disponível em: <<https://www.mdpi.com/1424-8220/20/23/6936>>. 25

 CAMALAN, S. et al. Change detection of amazonian alluvial gold mining using deep learning and sentinel-2 imagery. *Remote Sensing*, v. 14, n. 7, 2022. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/14/7/1746>>.

25

 DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

22, 23

 INSTRUTORGIS. QGIS: Satélite Amazonia-1 – Composição Colorida RGB. 2022. Disponível em: <<https://www.instrutorgis.com.br/qgis-satelite-amazonia1-composicao-colorida-rgb>>.

8

 KASELIMI, M. et al. A vision transformer model for convolution-free multilabel classification of satellite imagery in deforestation monitoring. *IEEE Transactions on Neural Networks and Learning Systems*, p. 1–9, 2022. ISSN 2162-2388.

25

 LIU, Z. et al. Swin transformer v2: Scaling up capacity and resolution. In: *CVPR 2022*. [s.n.], 2022. Disponível em: <<https://www.microsoft.com/en-us/research/publication/swin-transformer-v2-scaling-up-capacity-and-resolution/>>.

24

 MITCHELL, T. M. *Machine Learning*. New York: McGraw-Hill, 1997. ISBN 978-0-07-042807-2.

9

 PLANETCO, K. *Planet: Understanding the Amazon from Space - Use satellite data to track the human footprint in the Amazon rainforests*. Kaggle, 2016. Disponível em: <<https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/data>>.

27

 SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.*, v. 3, p. 210–229, 1959.

9

 WANG, D. et al. An empirical study of remote sensing pretraining. *arXiv preprint arXiv:2204.02825*, 2022.

25