

Evolução das Arquiteturas de Deep Learning na Classificação de Uso do Solo em Imagens de Satélite

Uma Comparação entre VGG, ResNet e Vision Transformers

CPE 727 - Aprendizado Profundo

Rafael Tadeu Cardoso dos Santos

December 11, 2025

Table of Contents

1 Introdução

- ▶ Introdução
- ▶ Modelos
- ▶ Metodologia
- ▶ Resultados
- ▶ Conclusão
- ▶ Referências Bibliográficas

Objetivo

1 Introdução

Este trabalho propõe um estudo comparativo entre três arquiteturas de Deep Learning aplicadas ao dataset **EuroSAT** [1][2].

Serão avaliadas:

- **MLP**: representando redes perceptron multicamadas simples como baseline de comparação;
- **VGG-16** [3]: representando redes convolucionais profundas tradicionais;
- **ResNet-50** [4]: representando redes residuais modernas
- **Vision Transformer (ViT-B/16)** [5]: representando o estado da arte baseado em mecanismos de atenção.

A ideia é entender como essas arquiteturas desempenham em problemas de visão computacional.

Dataset: EuroSAT

1 Introdução - Dados

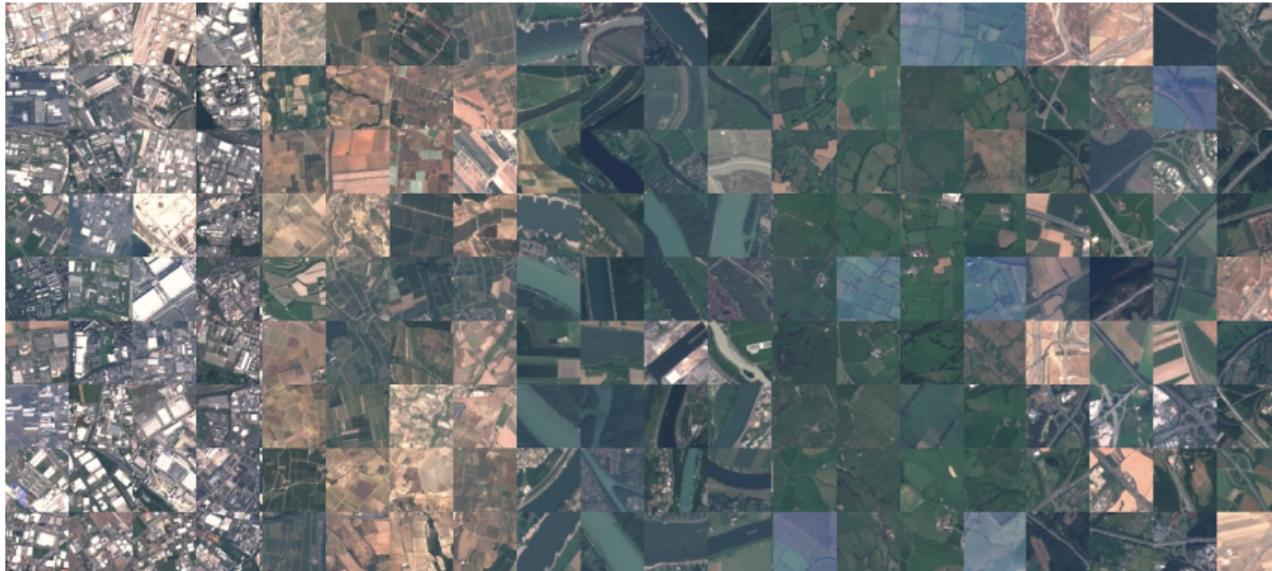


Figure: Figura adaptada do repositório original¹.

¹repositório

Dataset: EuroSAT

1 Introdução - Dados

Detalhes do dataset EuroSAT:

- Imagens de satélite Sentinel-2 da ESA
- 27.000 imagens RGB de 64x64 pixels
- 10 classes de uso do solo

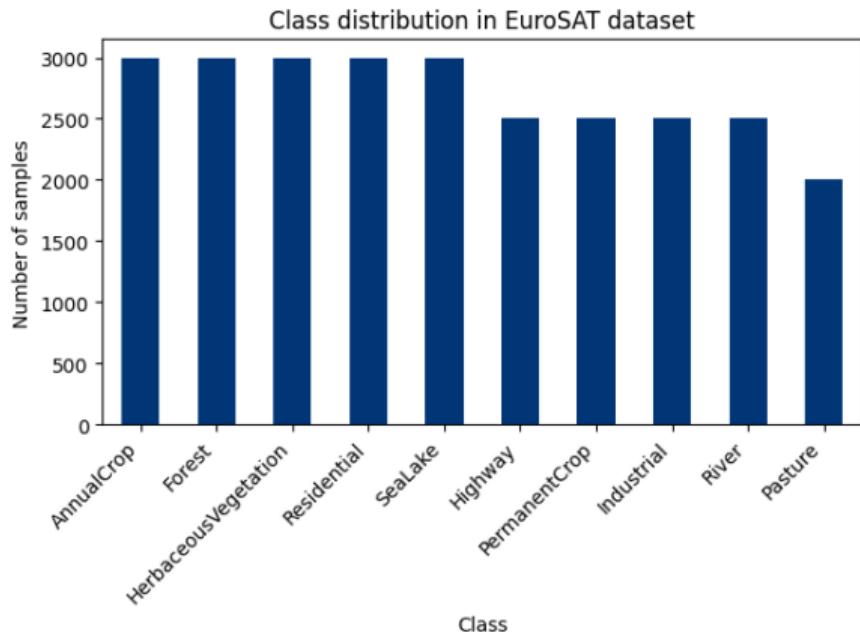


Table of Contents

2 Modelos

- ▶ Introdução
- ▶ Modelos
- ▶ Metodologia
- ▶ Resultados
- ▶ Conclusão
- ▶ Referências Bibliográficas

MLP (Baseline)

2 Modelos

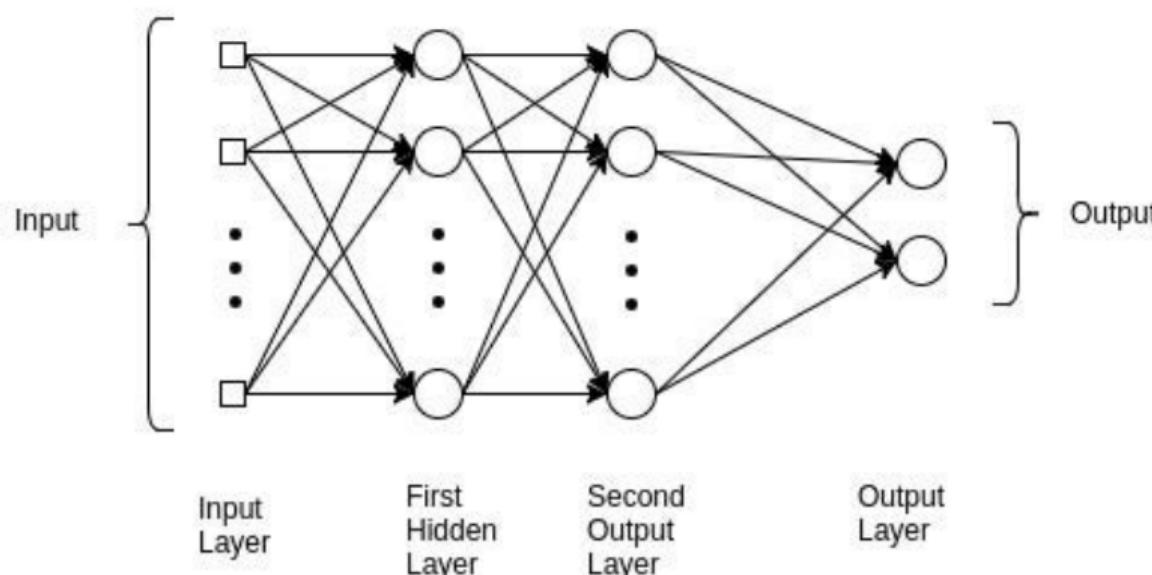
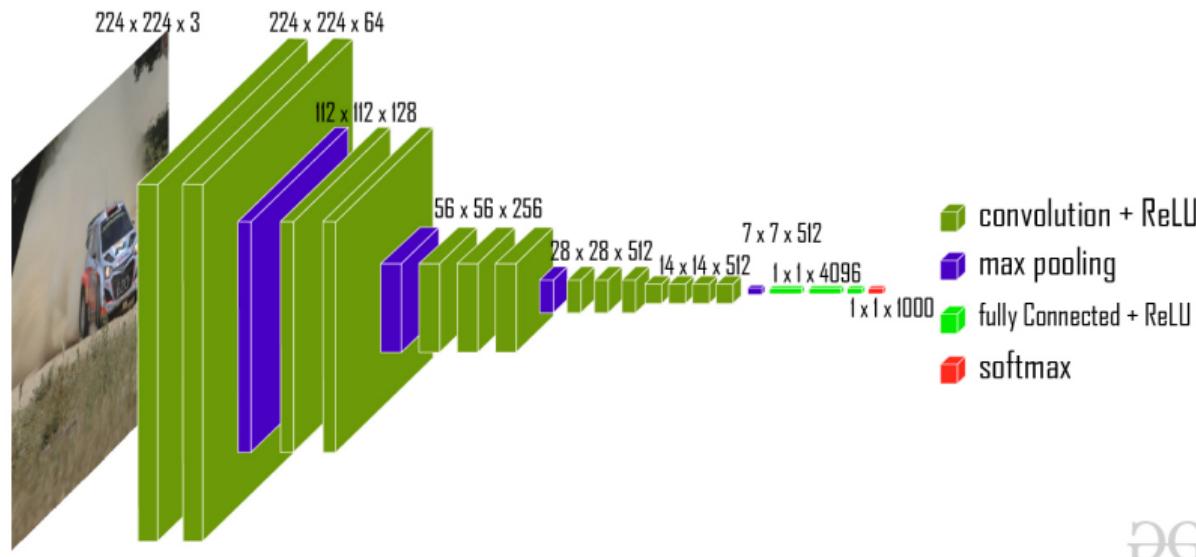


Figure: Arquitetura da MLP.

VGG-16

2 Modelos



DE

Figure: Arquitetura da VGG-16² [3].

²Figura retirada do site

ResNet-50

2 Modelos

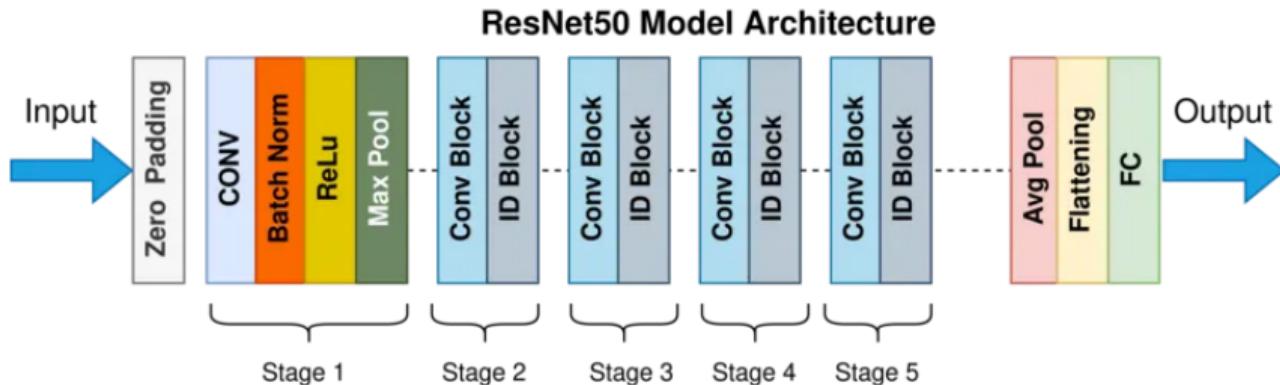


Figure: Arquitetura da ResNet-50 [4].

Vision Transformer (ViT-B/16)

2 Modelos

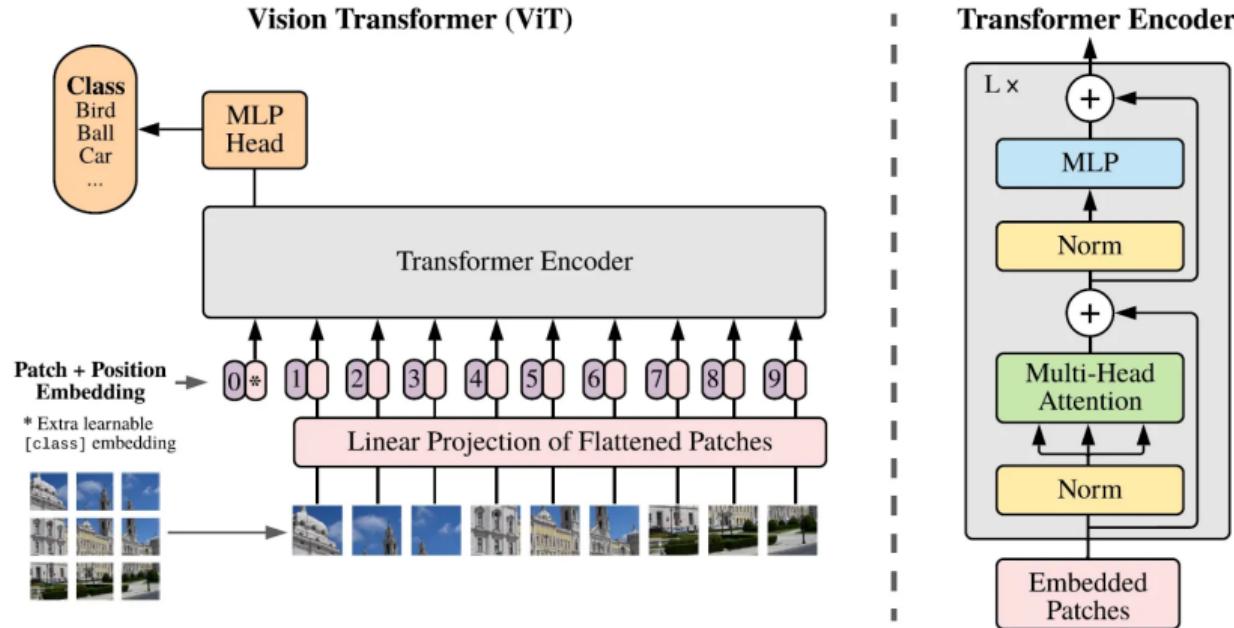


Figure: Arquitetura do Vision Transformer (ViT-B/16) [5].

Modificações nos Modelos Pré-treinados

2 Modelos

Nenhuma das arquiteturas pré-treinadas teve suas camadas convolucionais ou de atenção alteradas, apenas as camadas finais foram adaptadas para o problema específico de classificação do EuroSAT. Além disso, todos os modelos pré-treinados tiveram suas camadas base congeladas durante o treinamento inicial, focando o aprendizado nas camadas finais adaptadas.

- MLP: Nenhuma modificação necessária;
- VGG-16: Substituição da camada final do classificador fully connected (FC) para uma camada de Dropout10 e uma FC com 10 saídas (classes da EuroSAT) ;
- ResNet-50: Substituição da camada final fully connected para uma FC com 10 classes;
- ViT-B/16: Alteração da MLP Head para 10 classes de saída.

Table of Contents

3 Metodologia

- ▶ Introdução
- ▶ Modelos
- ▶ Metodologia
- ▶ Resultados
- ▶ Conclusão
- ▶ Referências Bibliográficas

Separação dos Dados

2 Metodologia - Treino, Validação e Teste

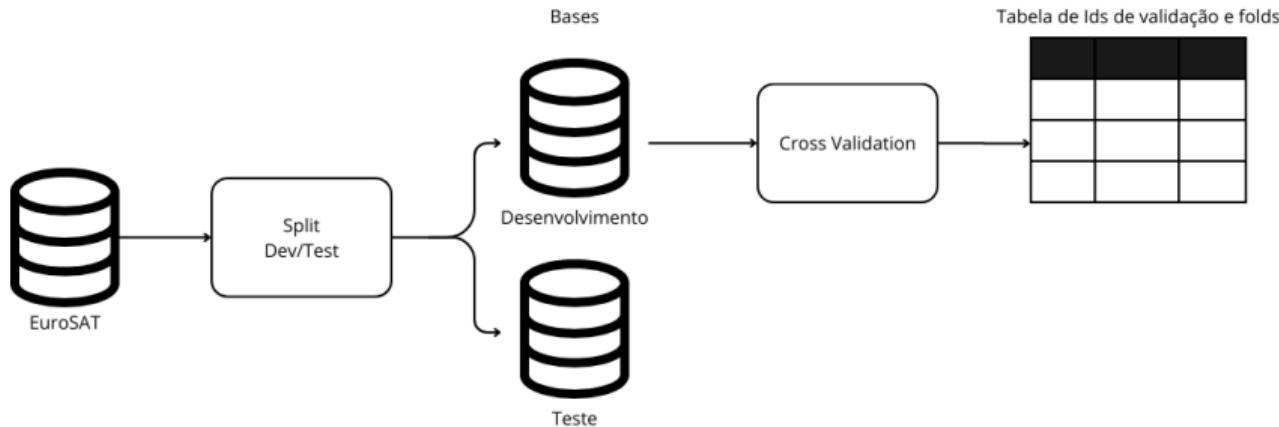


Figure: Processo de separação dos dados.

Separação dos Dados

2 Metodologia - Treino, Validação e Teste

- 80% dos dados para treino e validação (21.600 imagens);
- 20% dos dados para teste (5.400 imagens);
- Validação cruzada k-fold estratificado com k=5 para treino e validação.

Table: Distribuição de imagens por classe em cada fold

Classe	AC	For	HV	Hwy	Ind	Past	PC	Res	Riv	SL	Total
Treino	1895	1934	1911	1580	1612	1283	1596	1915	1602	1952	17280
Validação	474	484	477	395	404	321	398	479	401	487	4320

Separação dos Dados

2 Metodologia - Treino, Validação e Teste

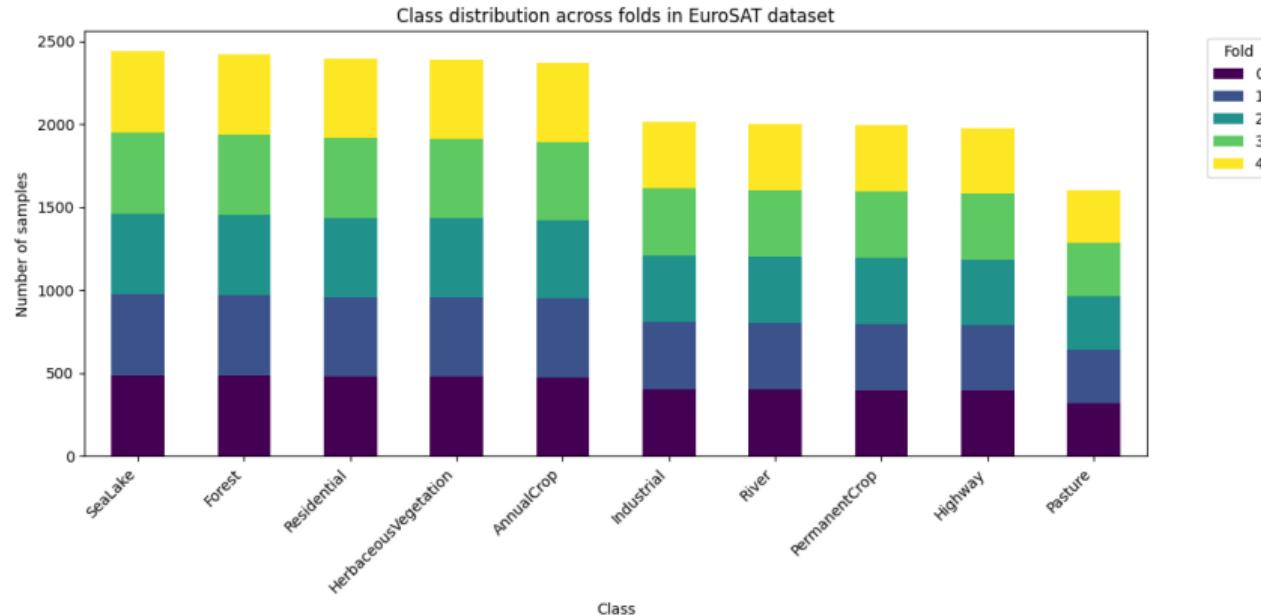


Figure: Distribuição de imagens por classe em cada fold.

Hiperparâmetros e Grid Search

2 Metodologia - Grid Search e Hiperparâmetros

Table: Configuração de hiperparâmetros para Grid Search

Modelo	Pré-treinado	Weights	Hidden Layers	Dropout	LR	Batch Size	Freeze BB
MLP	Não	-	[512, 256]	0.3	0.001	64	-
VGG-16	Sim	IMAGENET1K_V1	-	0.3	0.0001	64	Sim
VGG-16	Sim	IMAGENET1K_V1	-	0.5	0.0001	64	Sim
VGG-16	Sim	IMAGENET1K_V1	-	0.3	0.0001	128	Sim
VGG-16	Sim	IMAGENET1K_V1	-	0.5	0.0001	128	Sim
VGG-16	Sim	IMAGENET1K_V1	-	0.3	0.00001	64	Sim
VGG-16	Sim	IMAGENET1K_V1	-	0.5	0.00001	64	Sim
VGG-16	Sim	IMAGENET1K_V1	-	0.3	0.00001	128	Sim
VGG-16	Sim	IMAGENET1K_V1	-	0.5	0.00001	128	Sim
ResNet-50	Sim	DEFAULT	-	-	0.0001	64	Sim
ResNet-50	Sim	DEFAULT	-	-	0.0001	128	Sim
ResNet-50	Sim	DEFAULT	-	-	0.00001	64	Sim
ResNet-50	Sim	DEFAULT	-	-	0.00001	128	Sim
ViT-B/16	Sim	IMAGENET1K_V1	-	-	0.00001	64	Sim

LR: Learning Rate; Freeze BB: Freeze Backbone

Treinamento dos Modelos

2 Metodologia - Fluxo de Treinamento

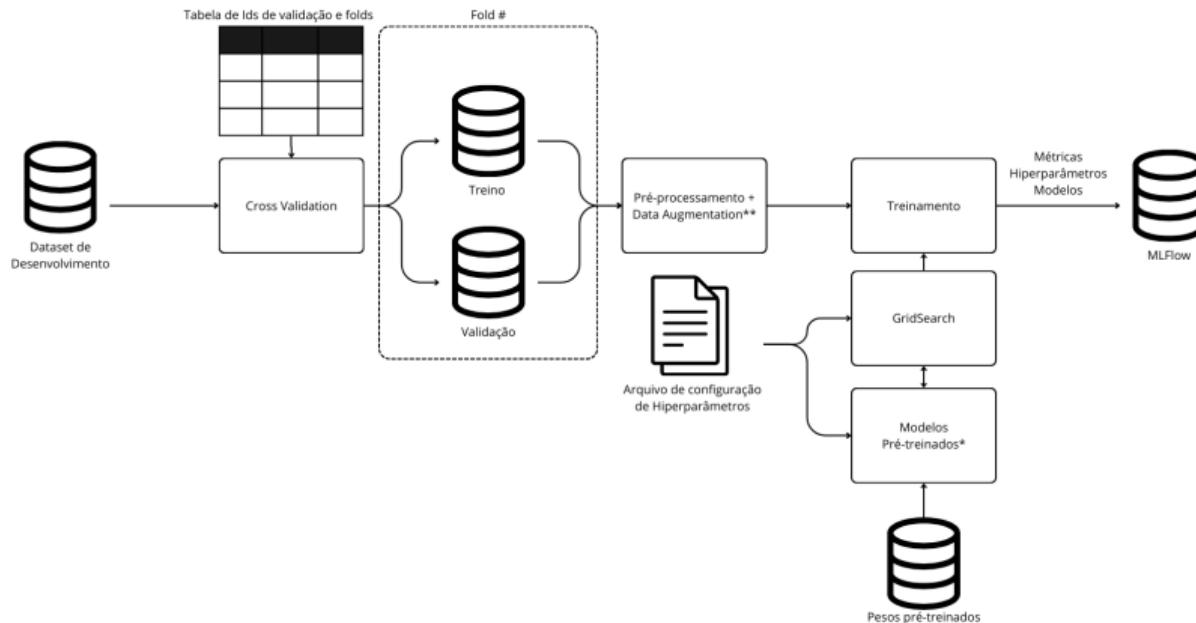


Figure: Processo de treinamento dos modelos. (*) Modelos pré-treinados na ImageNet, exceto MLP. (**) Data Augmentation usando RandomFlip, RandomRotation e ColorJitter.

Table of Contents

4 Resultados

- ▶ Introdução
- ▶ Modelos
- ▶ Metodologia
- ▶ Resultados
- ▶ Conclusão
- ▶ Referências Bibliográficas

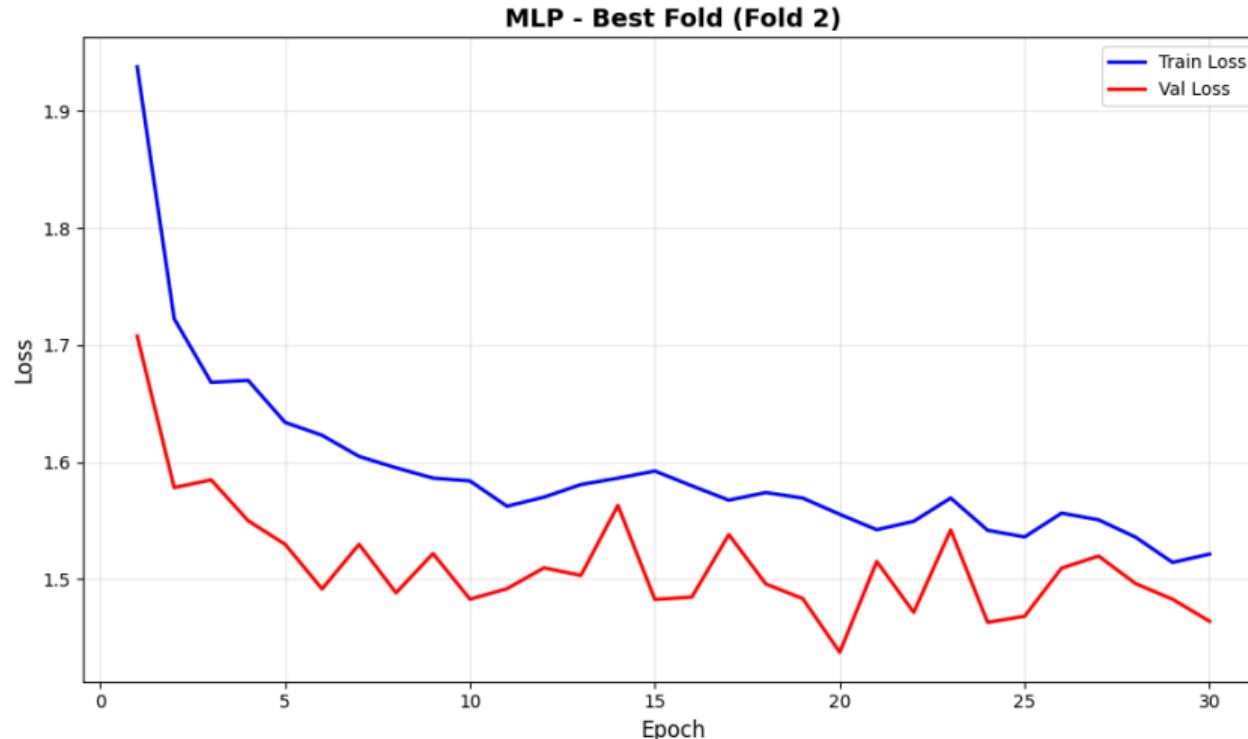
Resultados dos Treinamento em Validação Cruzada

4 Resultados

Model	Val Acc	Val Loss	Val F1	Val Prec	Val Rec
mlp	0.3336 ± 0.0226	1.7809 ± 0.0489	0.2758 ± 0.0261	0.3365 ± 0.0374	0.3336 ± 0.0226
resnet50	0.9163 ± 0.0019	0.2888 ± 0.0034	0.9161 ± 0.0020	0.9170 ± 0.0019	0.9163 ± 0.0019
resnet50	0.7933 ± 0.0043	0.9650 ± 0.0079	0.7914 ± 0.0038	0.8122 ± 0.0024	0.7933 ± 0.0043
resnet50	0.8281 ± 0.0011	0.7631 ± 0.0025	0.8261 ± 0.0012	0.8365 ± 0.0009	0.8281 ± 0.0011
resnet50	0.9068 ± 0.0012	0.3310 ± 0.0019	0.9066 ± 0.0013	0.9077 ± 0.0015	0.9068 ± 0.0012
vgg16	0.9008 ± 0.0075	0.3774 ± 0.0278	0.9003 ± 0.0075	0.9014 ± 0.0067	0.9008 ± 0.0075
vgg16	0.9062 ± 0.0049	0.2958 ± 0.0159	0.9058 ± 0.0049	0.9059 ± 0.0050	0.9062 ± 0.0049
vgg16	0.9748 ± 0.0031	0.0682 ± 0.0069	0.9748 ± 0.0031	0.9752 ± 0.0029	0.9748 ± 0.0031
vgg16	0.9779 ± 0.0014	0.0640 ± 0.0028	0.9779 ± 0.0014	0.9781 ± 0.0013	0.9779 ± 0.0014
vgg16	0.9064 ± 0.0054	0.3083 ± 0.0148	0.9060 ± 0.0053	0.9062 ± 0.0052	0.9064 ± 0.0054
vgg16	0.9004 ± 0.0057	0.3778 ± 0.0109	0.8999 ± 0.0054	0.9004 ± 0.0052	0.9004 ± 0.0057
vgg16	0.9044 ± 0.0063	0.2994 ± 0.0153	0.9040 ± 0.0063	0.9042 ± 0.0063	0.9044 ± 0.0063
vgg16	0.8975 ± 0.0068	0.4019 ± 0.0369	0.8973 ± 0.0068	0.8988 ± 0.0063	0.8975 ± 0.0068
vit_b_16	0.9294 ± 0.0015	0.2659 ± 0.0015	0.9295 ± 0.0015	0.9304 ± 0.0014	0.9294 ± 0.0015

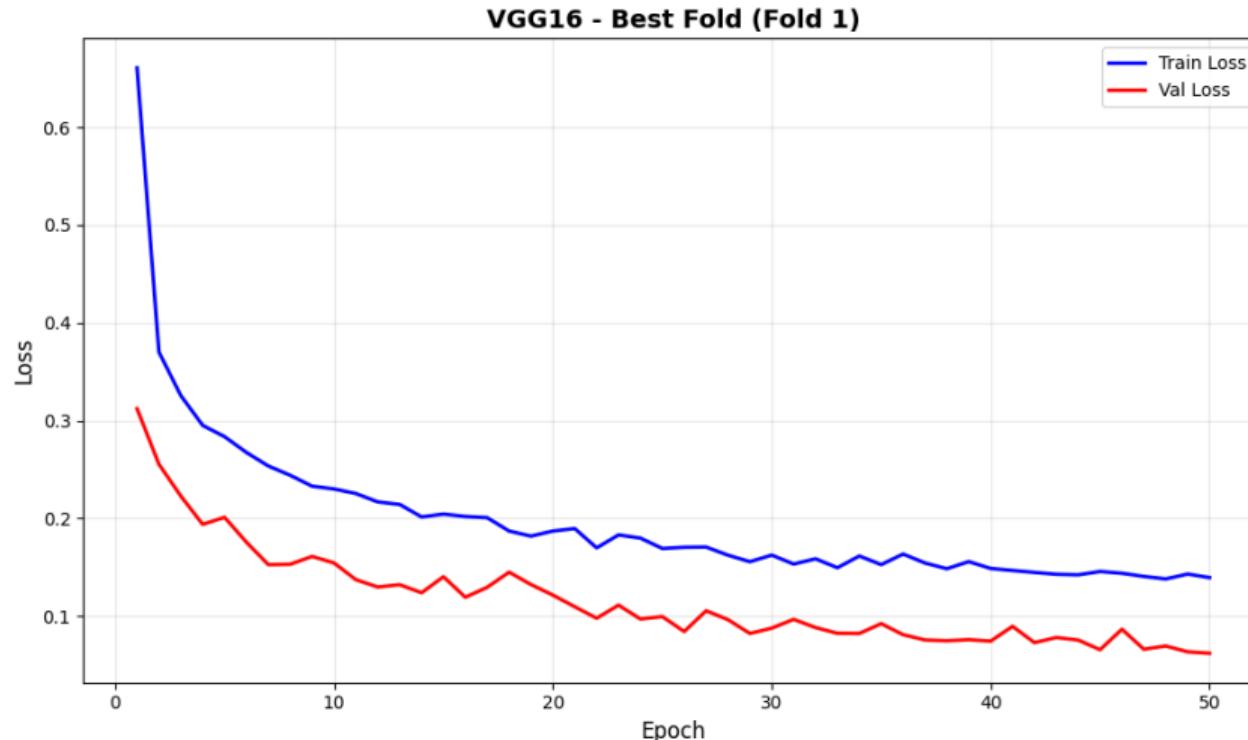
Resultados dos Treinamento em Validação Cruzada

4 Resultados - Curvas de Treinamento



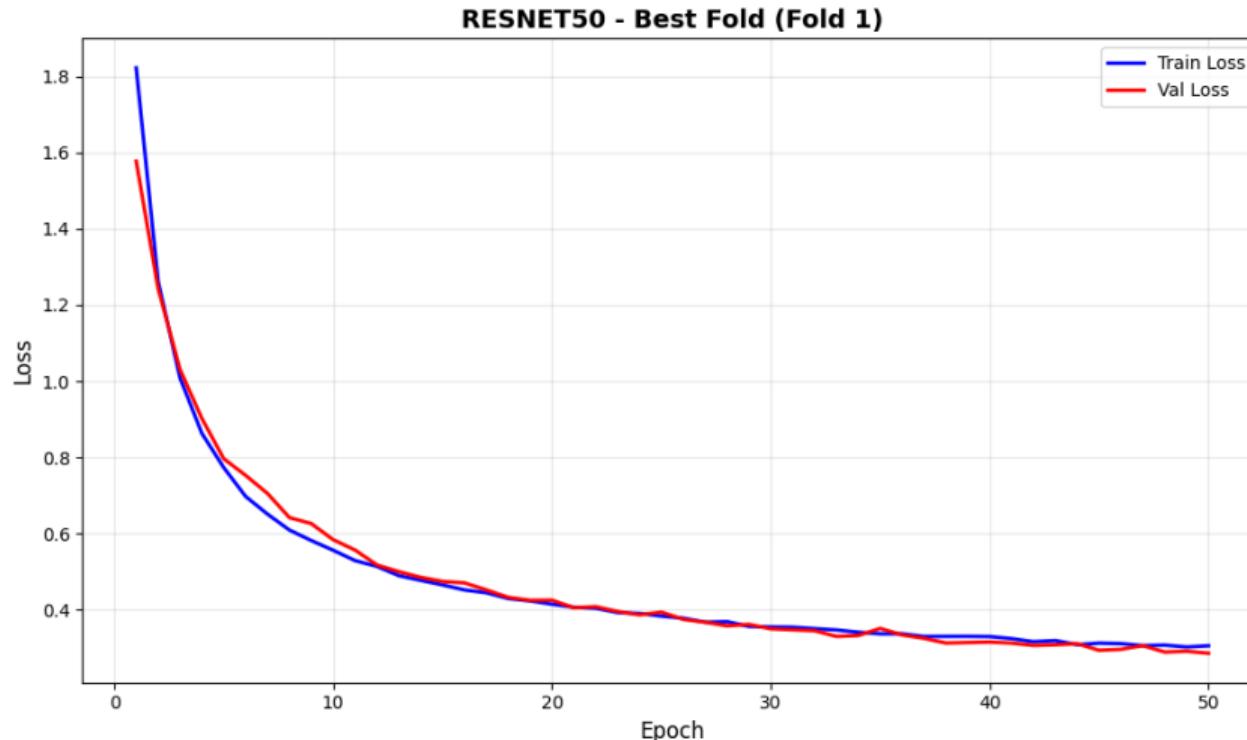
Resultados dos Treinamento em Validação Cruzada

4 Resultados - Curvas de Treinamento



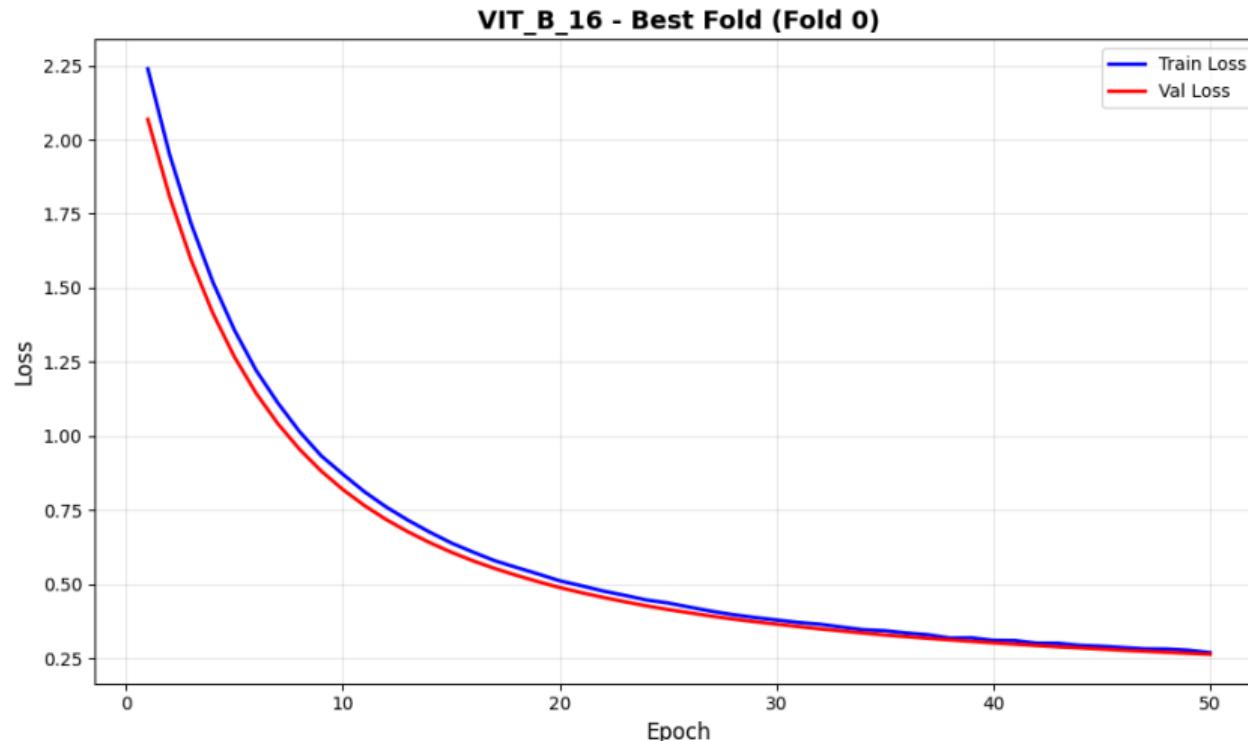
Resultados dos Treinamento em Validação Cruzada

4 Resultados - Curvas de Treinamento



Resultados dos Treinamento em Validação Cruzada

4 Resultados - Curvas de Treinamento



Resultados Finais no Conjunto de Teste

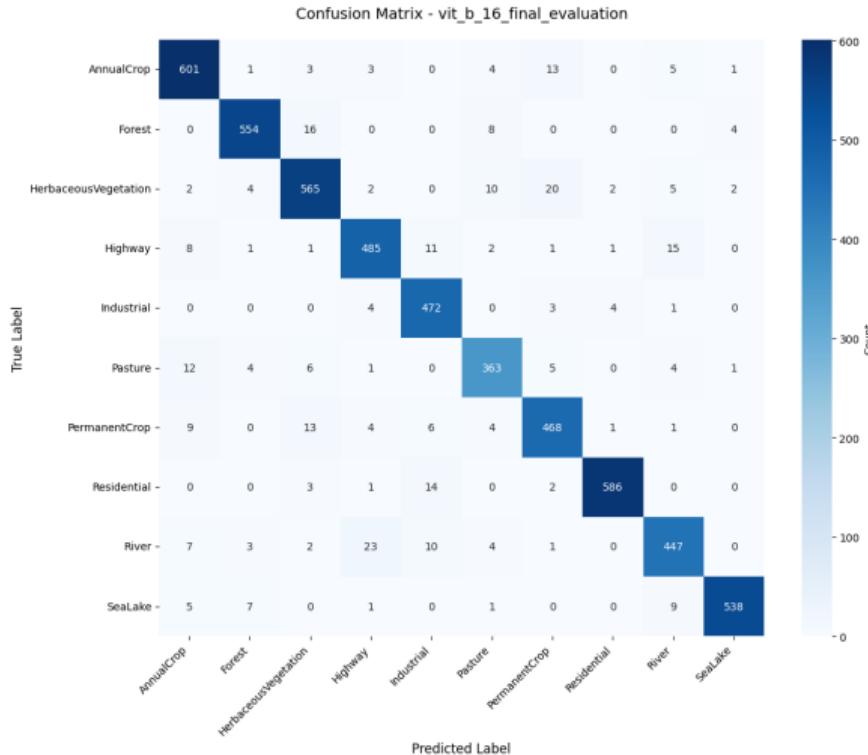
4 Resultados - Teste

Table: Acurárias Finais no Conjunto de Teste

Model	Accuracy
ViT-B/16	0.9405
VGG-16	0.9439
ResNet50	0.9131
MLP	0.2041

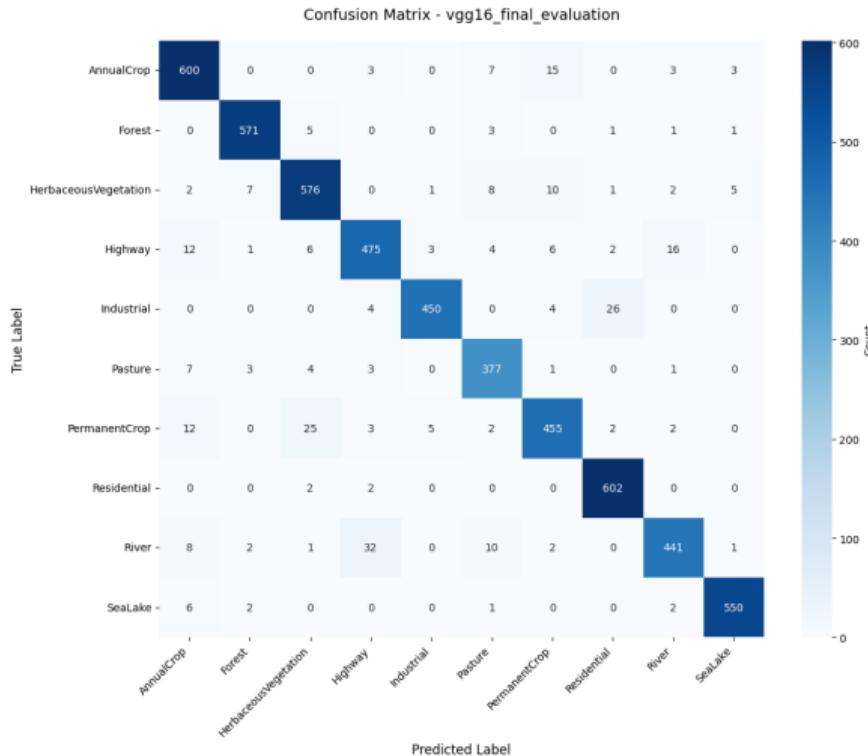
Resultados Finais no Conjunto de Teste

4 Resultados - Teste - ViT-B/16



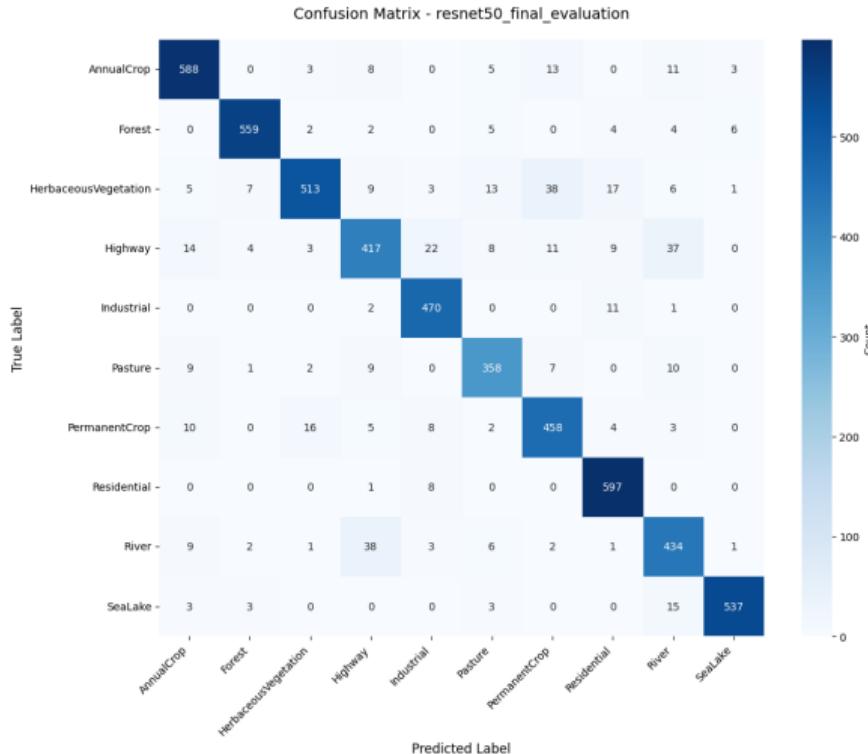
Resultados Finais no Conjunto de Teste

4 Resultados - Teste - VGG-16



Resultados Finais no Conjunto de Teste

4 Resultados - Teste - ResNet-50



Resultados Finais no Conjunto de Teste

4 Resultados - Teste - MLP

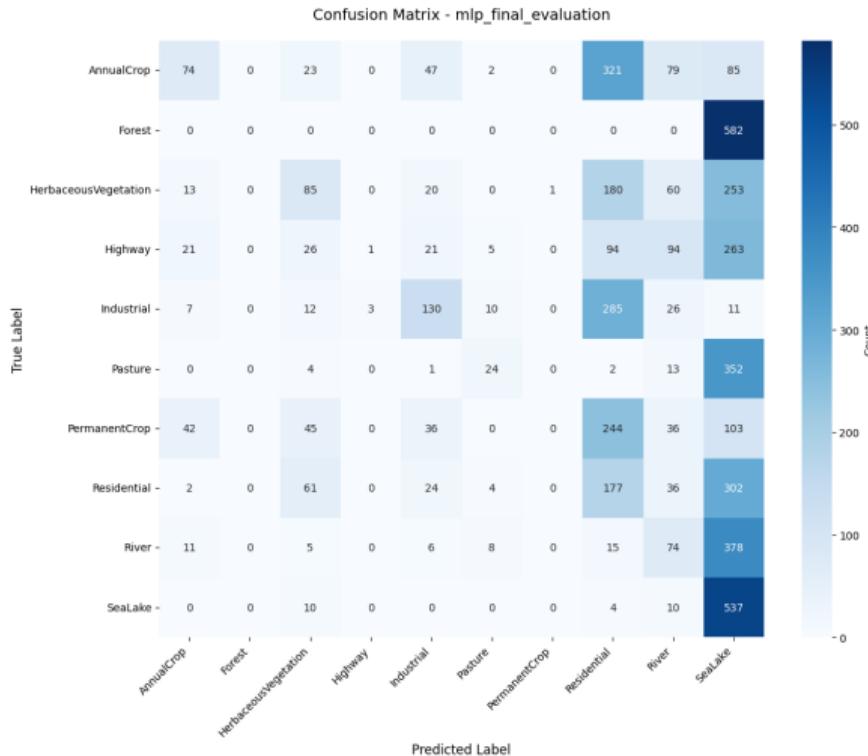


Table of Contents

5 Conclusão

- ▶ Introdução
- ▶ Modelos
- ▶ Metodologia
- ▶ Resultados
- ▶ Conclusão
- ▶ Referências Bibliográficas

Conclusão

5 Conclusão

- A VGG-16 obteve o melhor desempenho no conjunto de teste, seguida pelo ViT-B/16 e ResNet-50.
- A MLP teve um desempenho significativamente inferior, destacando a importância de arquiteturas mais complexas para tarefas de visão computacional.
- Redes convolucionais tradicionais ainda são altamente eficazes para classificação de imagens, apesar do avanço dos Transformers.
- Futuras pesquisas podem explorar outras arquiteturas e técnicas de pré-processamento para melhorar ainda mais o desempenho.

Table of Contents

6 Referências Bibliográficas

- ▶ Introdução
- ▶ Modelos
- ▶ Metodologia
- ▶ Resultados
- ▶ Conclusão
- ▶ Referências Bibliográficas

Referências Bibliográficas

6 Referências Bibliográficas

- [1] P. Helber, B. Bischke, A. Dengel, and D. Borth, “Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification,” in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 204–207, IEEE, 2018.
- [2] P. Helber, B. Bischke, A. Dengel, and D. Borth, “Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.
- [3] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.

Referências Bibliográficas

6 Referências Bibliográficas

- [4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.
- [5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021.

Evolução das Arquiteturas de Deep Learning na Classificação de Uso do Solo em Imagens de Satélite

Obrigado pela Atenção!

Alguma Pergunta?

Natanael Moura Junior

natmourajr@poli.ufrj.br, natmourajr@lps.ufrj.br