Aplicação de Patches Dinâmicos em Vision Transformers em Exames de Papanicolau

Vinícius H. Giovanini
Alexei Machado

Pontifícia Universidade Católica de Minas Gerais

13 de dezembro de 2024

PUC Minas

Sumário

PUC Minas

Análise de Imagens Médicas com DeepLearning

- Classificação de imagens médicas (radiografias, tomografias, ressonâncias magnéticas, etc.);
- Redes Neurais Convolucionais (CNNs);
- Visions Transformers (ViTs);
- Modelos Híbridos;
- Análise automática em grande escala;
- É essencial alcançar um elevado índice de acerto.

PUC Minas

O Exame de Papanicolau

- Exame de Papanicolau surgiu em 1928 com George Papanicolau;
- Detectar precocemente o câncer de colo de útero;
- Redução significativa na mortalidade por câncer cervical;
- Incidência de 4,86 casos por 100.000 mulheres no Brasil [Barcelos et al., 2017].

PUC Minas

Fluxo de Entrada e Saída do ViT

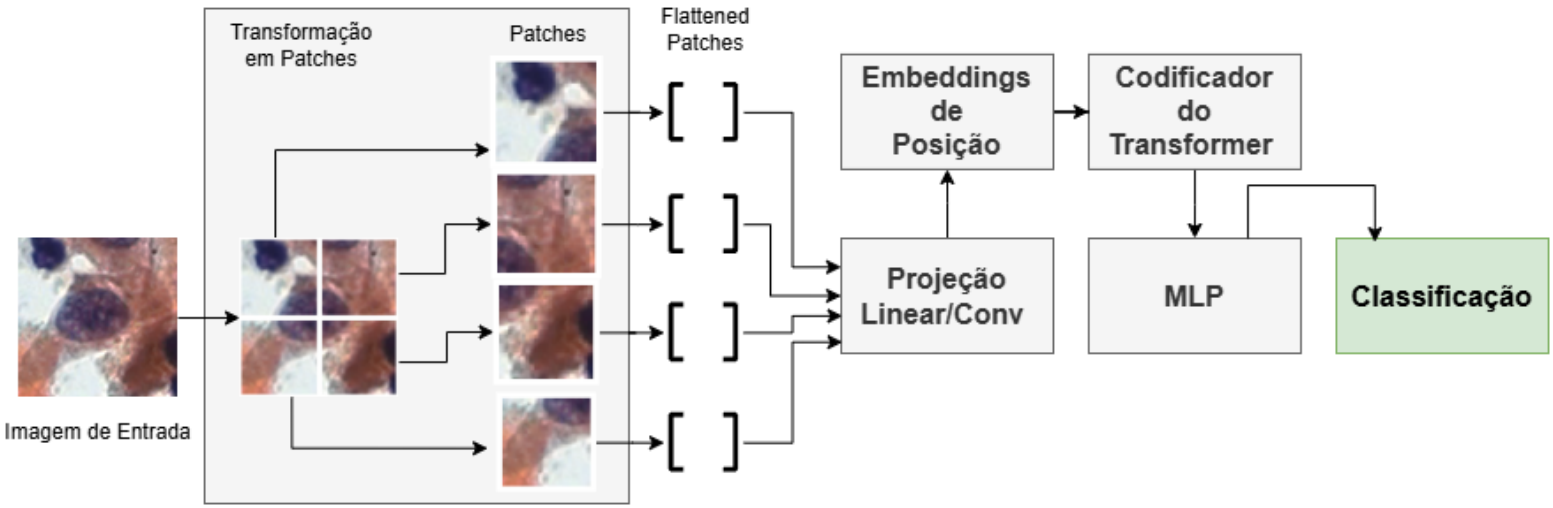


Figura: Fluxo de entrada do ViT

PUC Minas

Objetivos: ViT com Patches Dinâmicos

- Utilização do Vision Transformer;
- Base de Dados Centro de Reconhecimento e Inspeção de Células (CRIC) - UFOP
- Realização do Fine Tuning;
- Incorcoporação dos Patches Dinâmicos ao ViT;
  1. Seleção Randômica (SR);
  2. Seleção Randômica Aprimorada (RA);
  3. Seleção por Segmentação (SS);
- Comparação entre ViT e CNN com ensemble [N. Diniz et al., 2021].

PUC Minas

Artigos Relacionados

- Improving robustness for vision transformer with a simple dynamic scanning augmentation [Kotyan and Vargas, 2024]
- An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [Dosovitskiy et al., 2021]
- Visual Transformers: Token-based Image Representation and Processing for Computer Vision [Wu et al., 2021]
- Cric searchable image database as a public platform for conventional pap smear cytology data [Rezende et al., 2021]
- A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification [N. Diniz et al., 2021]

PUC Minas

Arquitetura do Vision Transformer (ViT)

- Introduzido em 2021 pelo Google Research;
- Baseado no Transformers (NLP);
- Arquitetura é composta:
  1. Divisão de imagens em patches;
  2. Adição do CLS Token e do Embeddings de Posição;
  3. Processamento pelo códificador (MLP e Auto-Atenção);
  4. MLP de Classificação.

PUC Minas

Arquitetura do Vision Transformer (ViT)

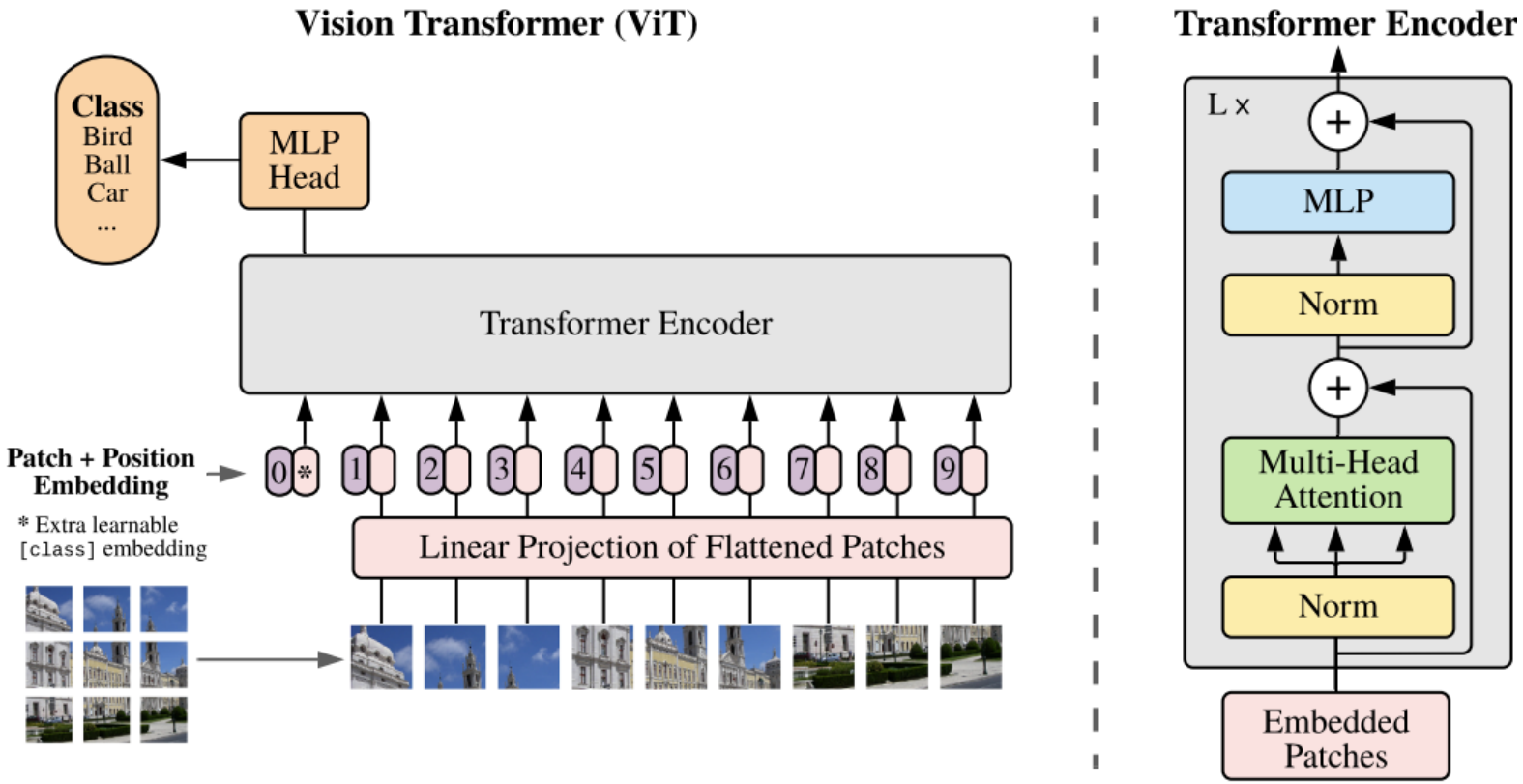


Figura: Arquitetura do ViT [Dosovitskiy et al., 2021]

PUC Minas

Patches

- Divisão de uma imagem em blocos menores de tamanho fixo;
- Cada patch é tratado como um token;
- Normalmente utiliza-se tamanho de 16x16 e 32x32;
- Os patches são mapeados para vetores;
- Menos recursos computacionais.

$$\text{Total de patches} = \left(\frac{H}{PS}\right) \times \left(\frac{W}{PS}\right)$$

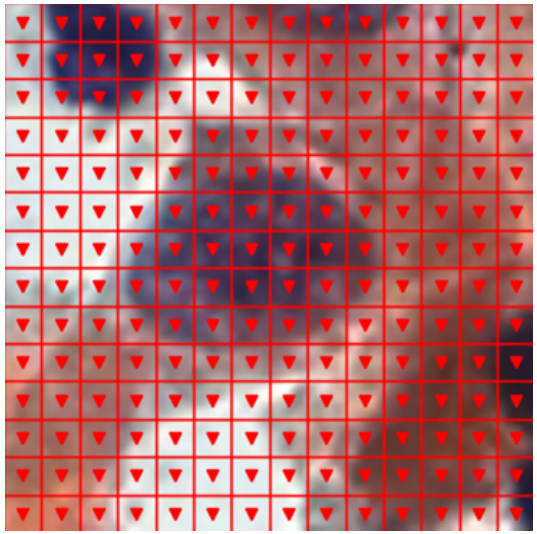


Figura: ASC-H 10

PUC Minas

Projeções

- **Projeção Linear**
  - Aplica-se uma camada Linear;
  - A imagem é dividida em pequenos patches, que são então achatados em vetores unidimensionais;
  - Utilizada no trabalho [Dosovitskiy et al., 2021];
  - Própria imagem.
- **Projeção Convolucional**
  - Filtro convolucional para gerar os embeddings;
  - Utilizada no trabalho [Wu et al., 2021];
  - Preserva-se informações espacial local;
  - Mapa de características.

PUC Minas

Materiais - Conjunto de Dados (Dataset)

- Centro de Reconhecimento e Inspeção de Células (CRIC);
- 400 imagens de exames de Papanicolau;
- Arquivo CSV e JSON;
- 11.534 núcleos de célula;
- Sistema Bethesda.
  1. Atypical squamous cells of undetermined signifcance (ASC-US)
  2. Atypical squamous cells cannot exclude a high-grade lesion (ASC-H)
  3. High-grade squamous intraepithelial lesion (HSIL)
  4. Low-grade squamous intraepithelial lesion (LSIL)
  5. Negative for intraepithelial lesion (NFIL)
  6. Squamous cell carcinoma (SCC)

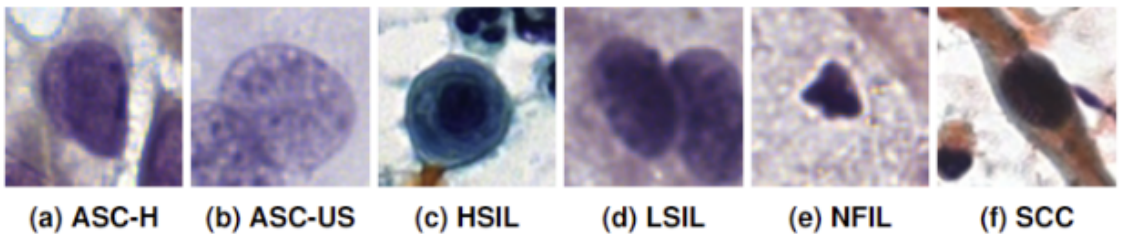


Figura: Recorte das células 90x90

PUC Minas

Balanceamento de dados

- Grande desbalanceamento entre os dados;
- 80/20 e 20% do Treino para o Teste;
- Albumentations;
- Corte Aleatório, Inversão Horizontal e Rotação;
- Balanceamento para 1000 imagens no Treino.

Divisão	**ASC-H**	**ASC-US**	**HSIL**	**LSIL**	**NFIL**	**SCC**
Treino (A/D)	592/1000	388/1000	1.090/1000	871/1000	4.339/1000	103/1000
Validação	185	122	341	272	1.356	33
Teste	148	96	272	217	1.084	25

Tabela: Quantidade de Imagens antes e depois do balanceamento

PUC Minas

Método - Introdução aos Patches Dinâmicos

- Explorar a sobreposição de patches;
- Extrair área de interesse;
- Substituição da classe tradicional de Embeddings por uma classe customizada;
- Introdução ao conceito de centros.

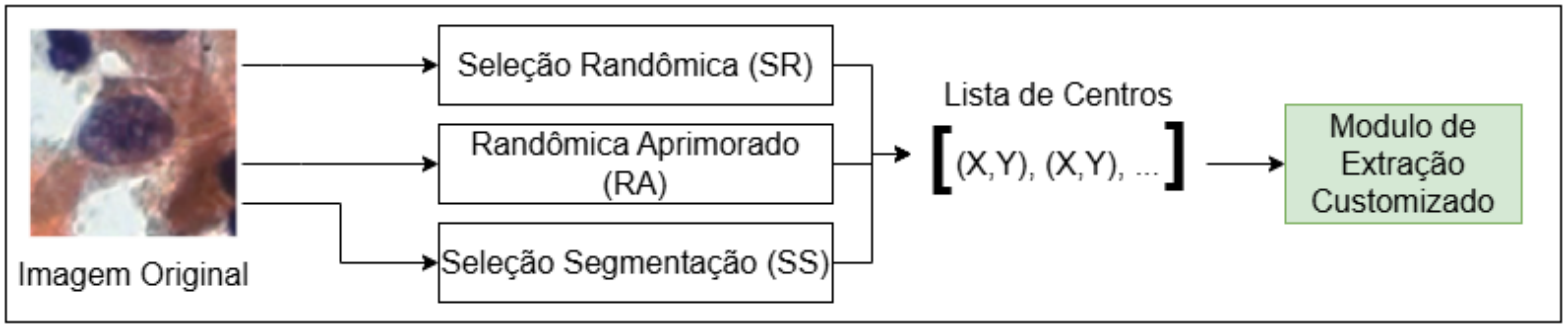


Figura: Fluxo Inicial de Extração de Patches com Centros

PUC Minas

Método - Seleção Randômica (SR)

- Método de extração completamente aleatório;
- Seleciona X centros a partir da quantidade total de patches necessários;
- Método rápido, possível utiliza-lo durante o treinamento;
- Sobreposição descontrolada.

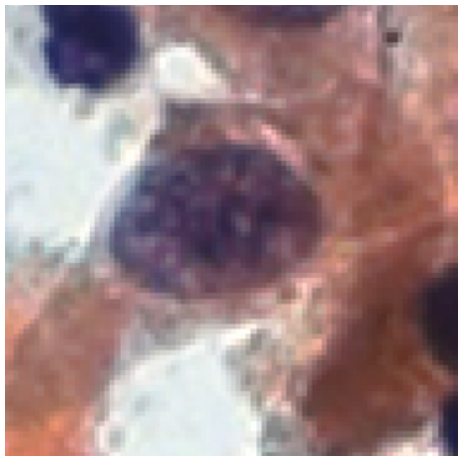
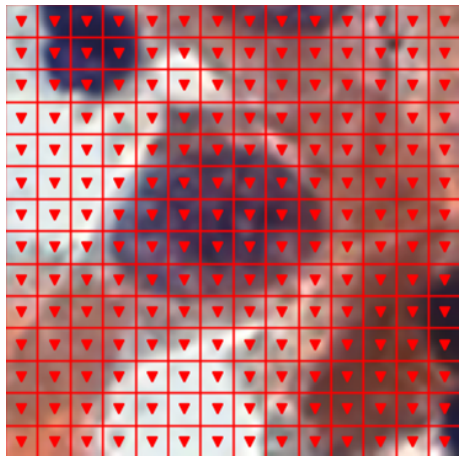
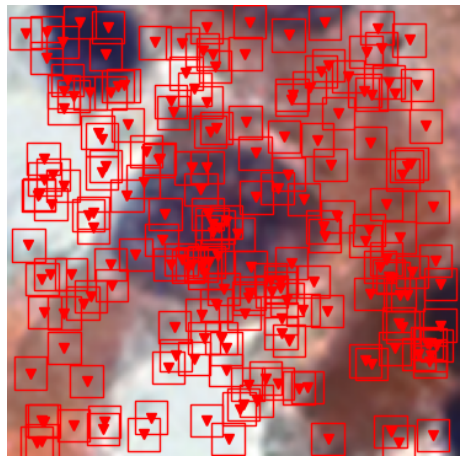


Imagem Original



Extração por Grid



Extração por SR

PUC Minas

Método - Randômico Aprimorado

- Método de extração aleatório;
- Algoritmo que controla sobreposição;
- Centro só é validado se não estiver dentro da área de outro patch já existente;
- Método custoso, não é realizado durante o treinamento;
- Sobreposição controlada.

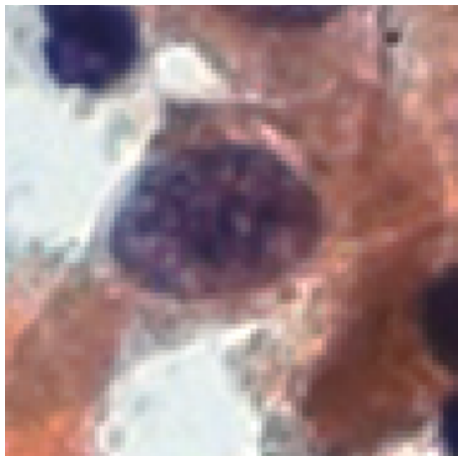
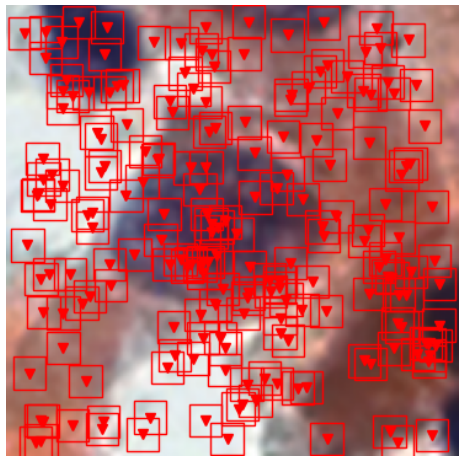
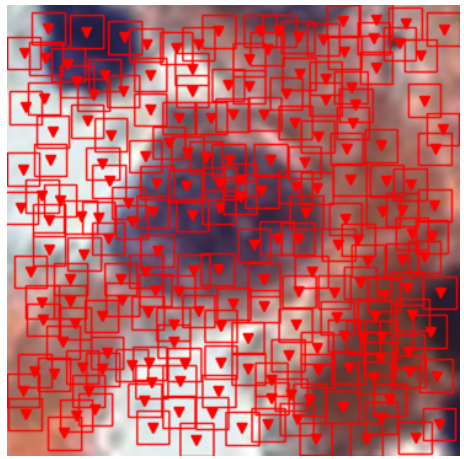


Imagem Original



Extração por SR



Extração por RA

PUC Minas

Método - Seleção por Segmentação

- Método de extração por área de interesse;
- Utiliza-se o GrabCut;
- Extrai os patches em Grid da máscara;
- Patches que sobram são selecionados randomicamente;
- Método custoso, não é realizado durante o treinamento;
- Sobreposição pouco controlada.

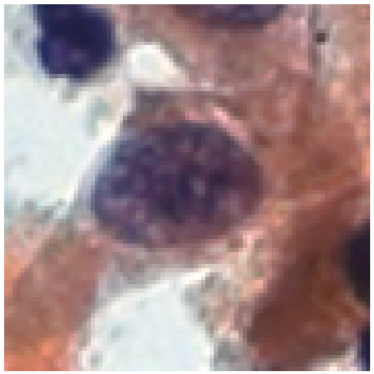


Imagem Original

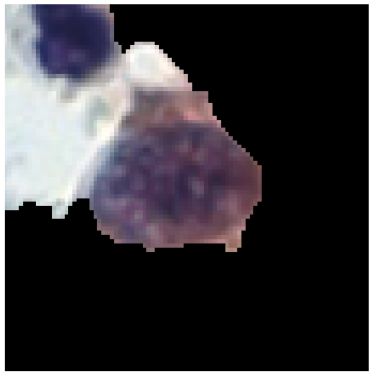


Imagem Recortada

Máscara

SS

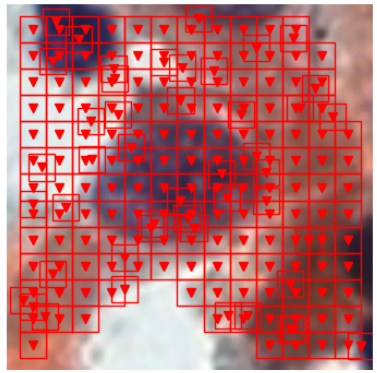
PUC Minas

Método - Seleção por Segmentação

Imagem Original

Imagem Recortada

Máscara



SS

PUC Minas

Método - Pré-processamento dos Patches Dinâmicos

- Pré-processamento para cada abordagem;
- A lista de centros foi salva em um dicionário;
- Classe personalizada de carregamento de dados (ImageFolder);
- Imagem, Rótulo, Nome da Imagem;
- Busca em um dicionário;
- Tempo do treinamento caiu para 3 horas em todos as extrações dinâmicas.

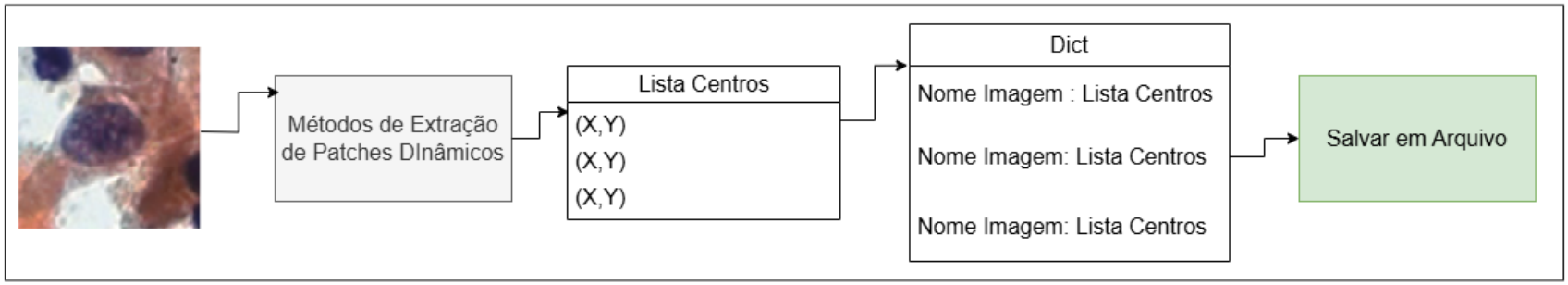


Figura: Fluxo de Pré-Processamento dos Centros

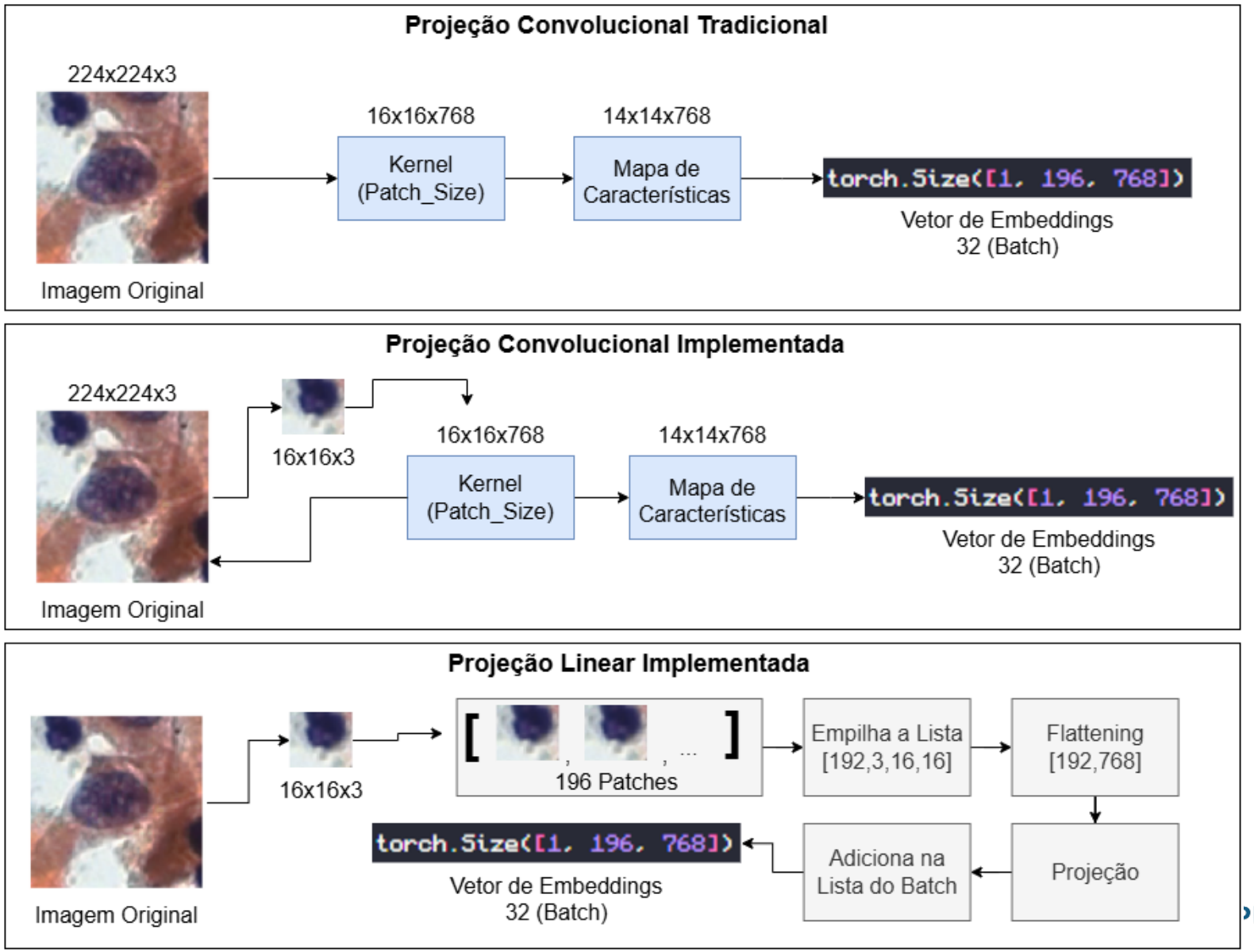
PUC Minas

Método - Modelos pré-treinados de ViT

- Google, Facebok (Meta), Microsoft;
- Foi utilizado os modelos do Google;
  - ViT-B/16;
  - ViT-S/16;
  - ViT-T/16;
- Disponibilizados no GitHub do Google Research JAX\FLEX;
- Hugging Face em Pytorch.

PUC Minas

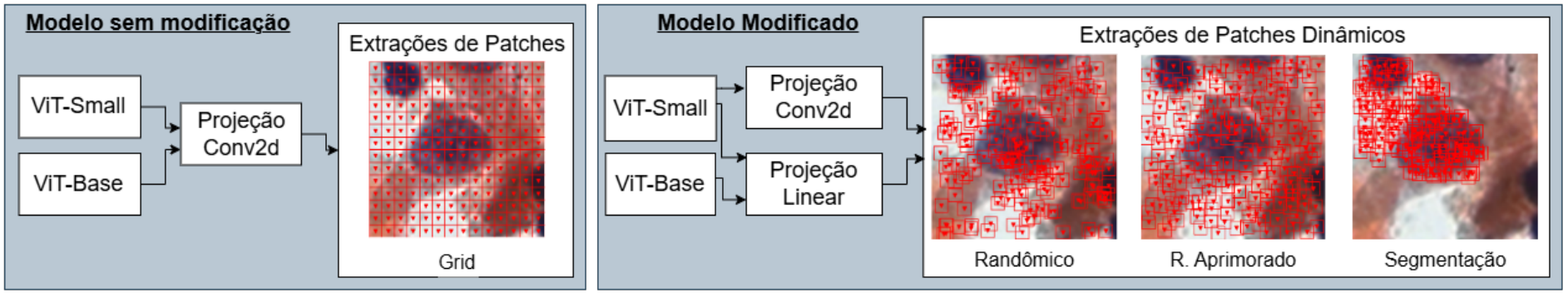
Método - Implementação das Projeções



PUC Minas

Fluxo de Experimentos

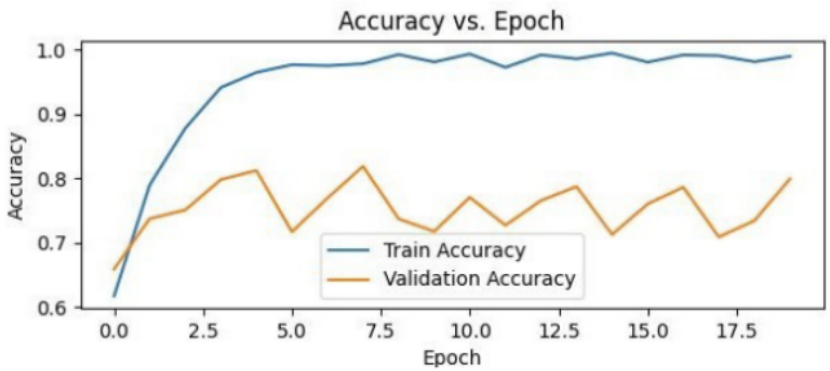
- 2 treinamentos do modelo com extração por Grid;
  - Projeção Convolucional;
  - Extração por Grid;
- 9 treinamentos do modelo com extração Dinâmica;
  - Projeção Convolucional;
  - Projeção Linear.



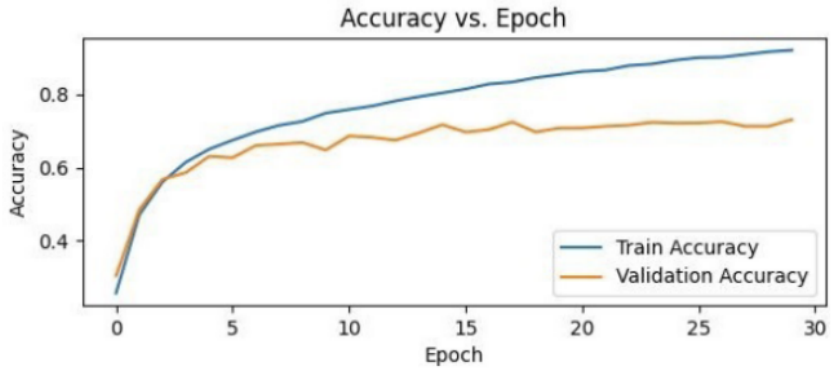
PUC Minas

Experimentos - Hiperparâmetros

- Taxa de Aprendizado (1e-5);
- Dropout (0.4);
- Weight decay (Descartado);
- Batch size (32);
- Épocas (60);
- Modelo ViT-Tiny foi descartado.



(a) Taxa de Aprendizado 1e-4



(b) Taxa de Aprendizado 1e-5

Figura: Alteração na Taxa de Aprendizado

PUC Minas

Experimentos - Finetuning - Extração por Grid

- Modelo Base tem 87.6 milhões de parâmetros;
- Modelo Small tem 22.1 milhões de parâmetros;
- Descongelou-se somente o último MLP do último bloco do modelo ViT-Base;
- Descongelou-se 5 blocos completos no modelo ViT-Small.

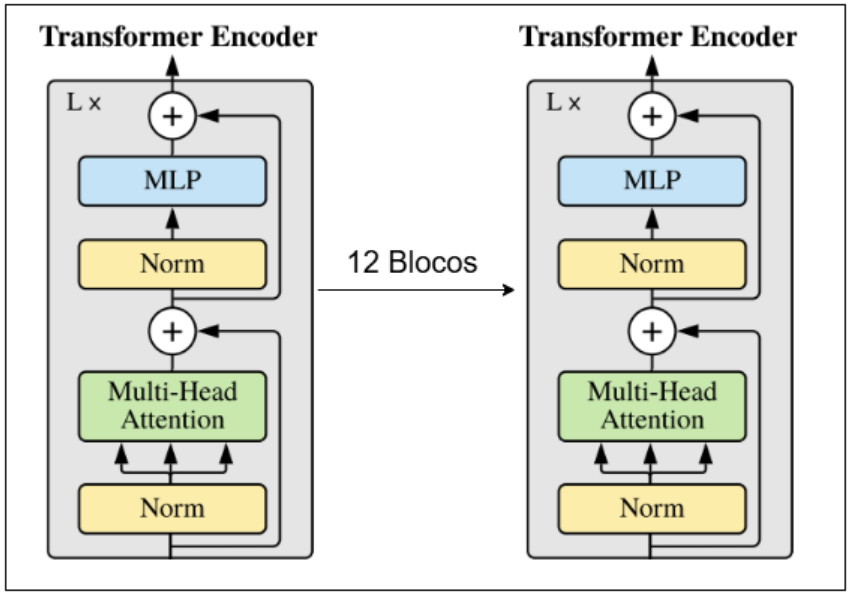
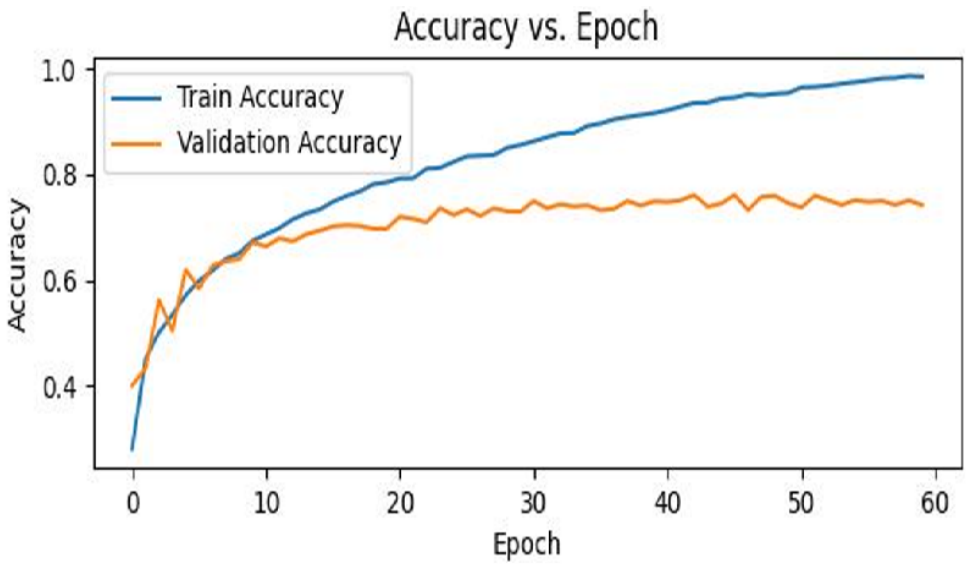


Figura: Bloco do Encoder - ViT

PUC Minas

Gráficos ViT-Base

- 76% (Validação), 75% (Teste)



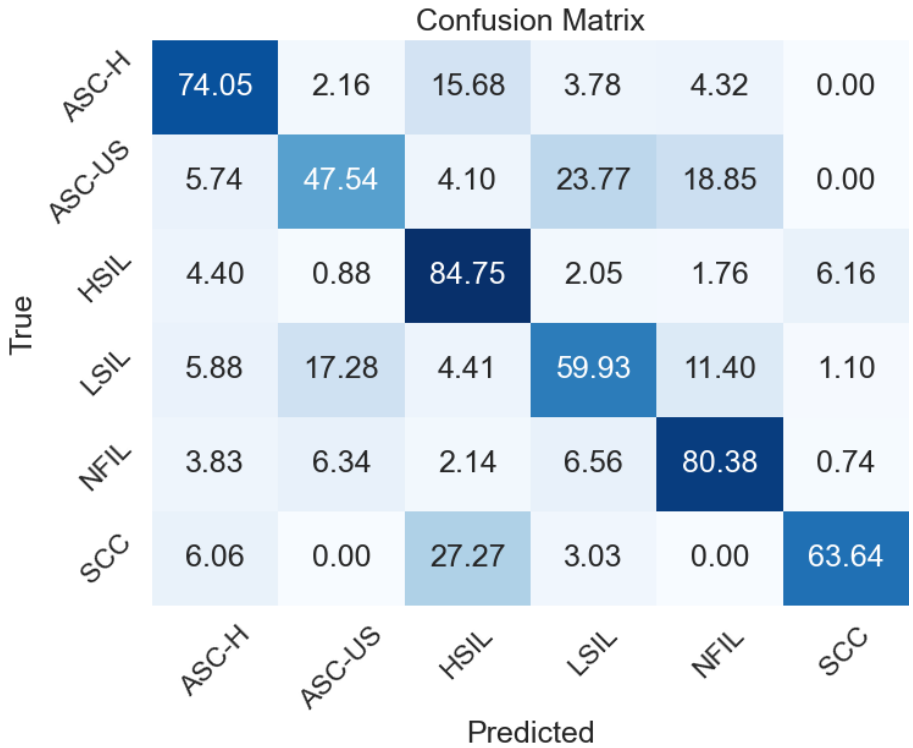
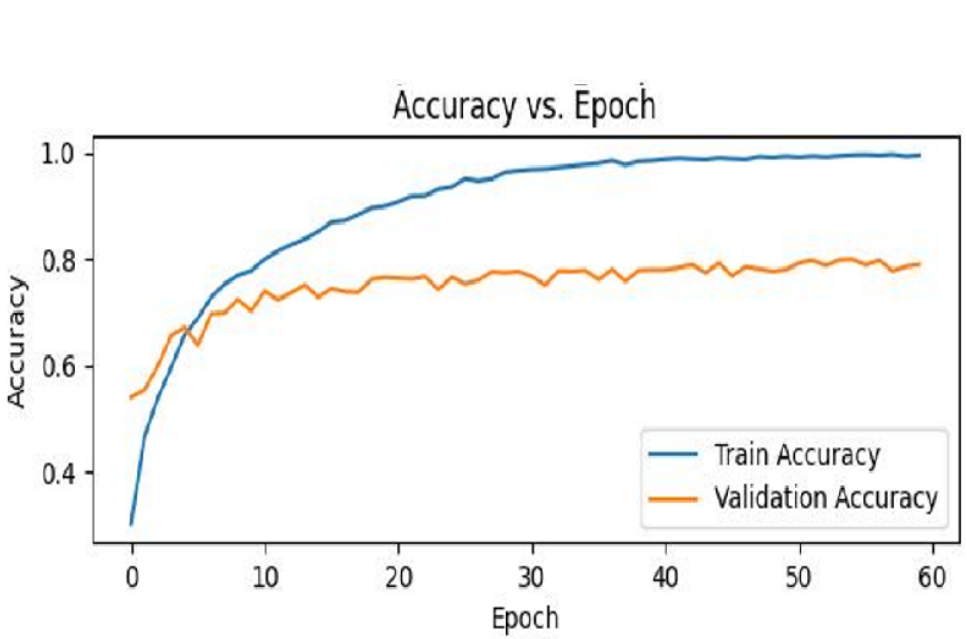


Figura: Modelo ViT-Base

PUC Minas

Gráficos ViT-Small

- 79% (Validação), 81% (Teste)



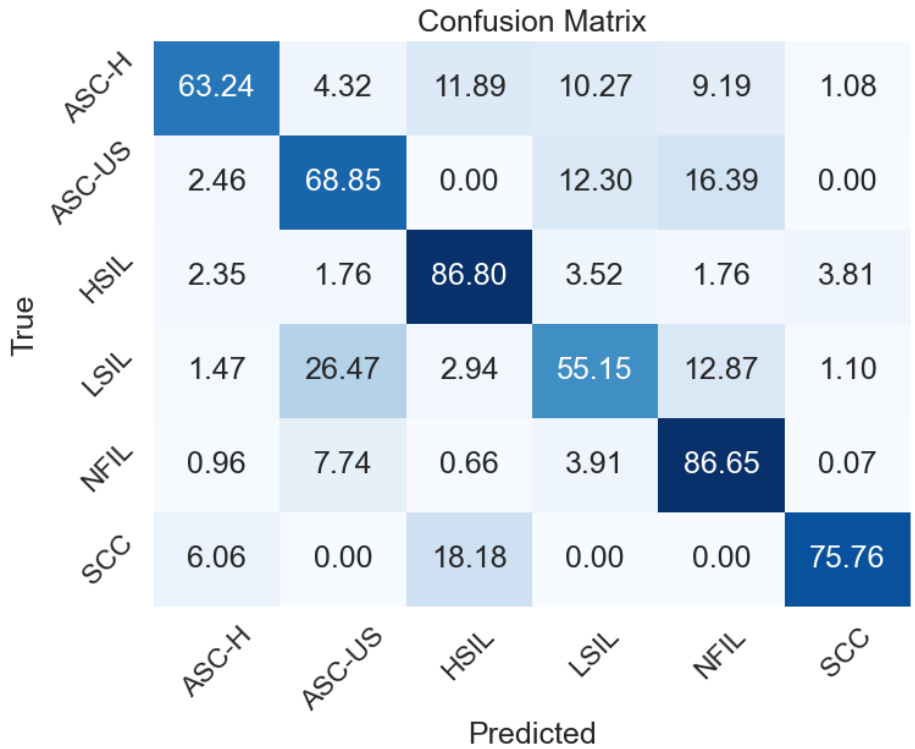
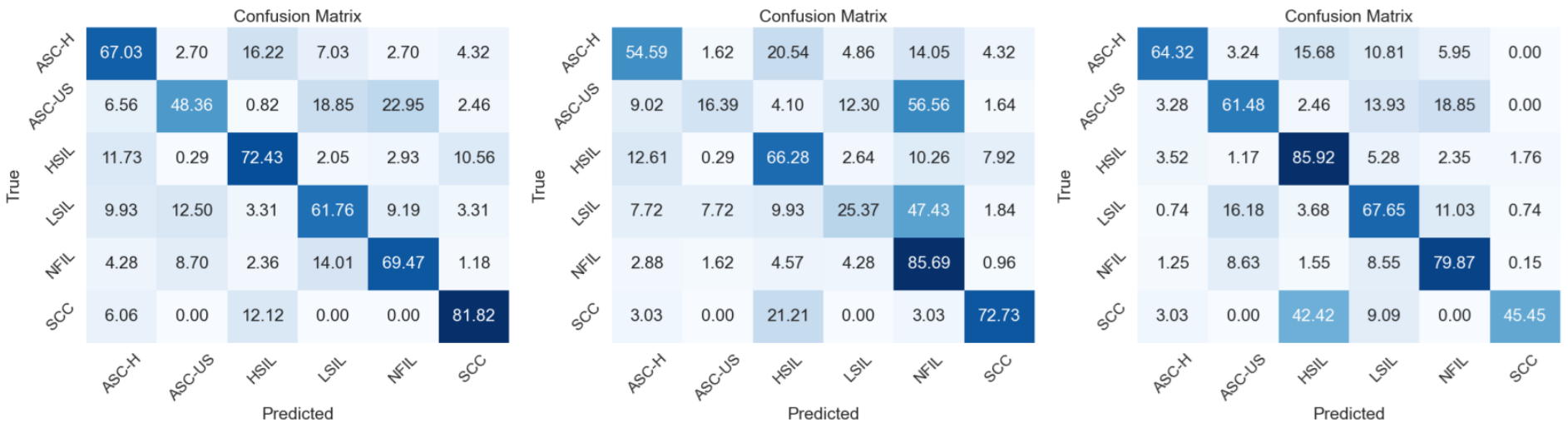


Figura: Modelo ViT-Small

PUC Minas

Experimentos - FineTuning - Extração Dinâmica

- É repetindo as mesmas arquiteturas e hiperparâmetros, para o ViT-Small e ViT-Base;
- Abordagem de extração de patches dinâmica;
- Tipo de projeção Linear e Convolucional;
- Comparando a acurácia do modelo de extração por Grid.



(a) SR Base Linear **(b) SS Small Linear** **(c) RA Small Conv**

Figura: Matriz de Confusão - Melhores Resultados Dinâmicos

PUC Minas

Resultados

- Método CNN com Ensemble obteve acurácia de 95%.

ViT	Projeção	Seleção Randômica	Randômica Aprimorada	Seleção por Segmentação
Base	Linear	69.54	64.92	59.50
Small	Linear	63.73	70.14	70.73
Small	Conv2d	75.02	77.30	75.78

Tabela: Acurácia dos Resultados Finais - Extração de Patches Dinâmicos

PUC Minas

Análise dos Resultados

- Nenhuma abordagem de extração, superou a CNN com ensemble;
- A extração de patches dinâmicos impactou a acurácia;
- A projeção Convolucional apresenta vantagens em relação a Linear;
- O melhor método dinâmico foi o Randômico Aprimorado com Projeção Convolucional (77.3% acc);
- Esse conjunto de dados favorece a sobreposição exagerada de patches, na extração por SS.

PUC Minas

Conclusão e Trabalhos Futuros

- Extração Randômica Aprimorada demonstrou melhor resultado entre os métodos dinâmicos;
- O método RA com projeção convolucional utilizando o modelo ViT-Base, com mais camadas descongeladas, mostra-se promissor;
- Melhoria do balanceamento através de Redes Generativas (GANs);
- Outros métodos de extração de patches, que seguem a visão humana;
- Treinamento de mais camadas do modelo ViT-Base.

PUC Minas

Referências I

Barcelos, M. R. B., Lima, R. d. C. D., Tomasi, E., Nunes, B. P., Duro, S. M. S., and Facchini, L. A. (2017).
Quality of cervical cancer screening in brazil: external assessment of the pmaq.
REV SAUDE PUBL, 51:67.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021).
An image is worth 16x16 words: Transformers for image recognition at scale.
ArXiv, abs/2010.11929.

Kotyan, S. and Vargas, D. V. (2024).
Improving robustness for vision transformer with a simple dynamic scanning augmentation.
Neurocomputing, 565:127000.

N. Diniz, D., T. Rezende, M., G. C. Bianchi, A., M. Carneiro, C., J. S. Luz, E., J. P. Moreira, G., M. Ushizima, D., N. S. de Medeiros, F., and J. F. Souza, M. (2021).
A deep learning ensemble method to assist cytopathologists in pap test image classification.
J IMAGING SCI, 7(7).

Rezende, M. T., Silva, R., Bernardo, F. d. O., Tobias, A. H. G., Oliveira, P. H. C., Machado, T. M., Costa, C. S., Medeiros, F. N. S., Ushizima, D. M., Carneiro, C. M., and Bianchi, A. G. C. (2021).
Cric searchable image database as a public platform for conventional pap smear cytology data.
Scientific Data, 8(1):151.

Wu, B., Xu, C., Dai, X., Wan, A., Zhang, P., Tomizuka, M., Keutzer, K., and Vajda, P. (2021).
Visual transformers: Token-based image representation and processing for computer vision.
ArXiv, abs/2006.03677.

PUC Minas

Obrigado!

vgiovanini@sga.pucminas.br
alexeimcmachado@gmail.com

PUC Minas

PUC Minas