



Transfer Learning para detecção de cromossomos

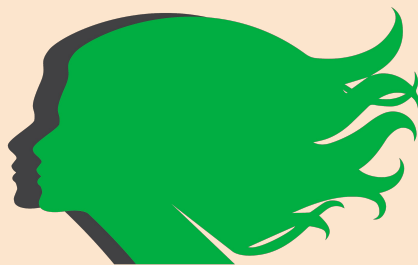
Maria Fernanda S. Andrade

Sobre mim :)

- ▶ @NaanAdr
- ▶ fsandrade25@gmail.com



pyladies
Recife



WOMEN IN DATA SCIENCE
RECIFE



PUGPE

PYTHON USER GROUP - PERNAMBUCO

Agradecimentos



CRCN|NE

CENTRO REGIONAL DE CIÊNCIAS
NUCLEARES DO NORDESTE

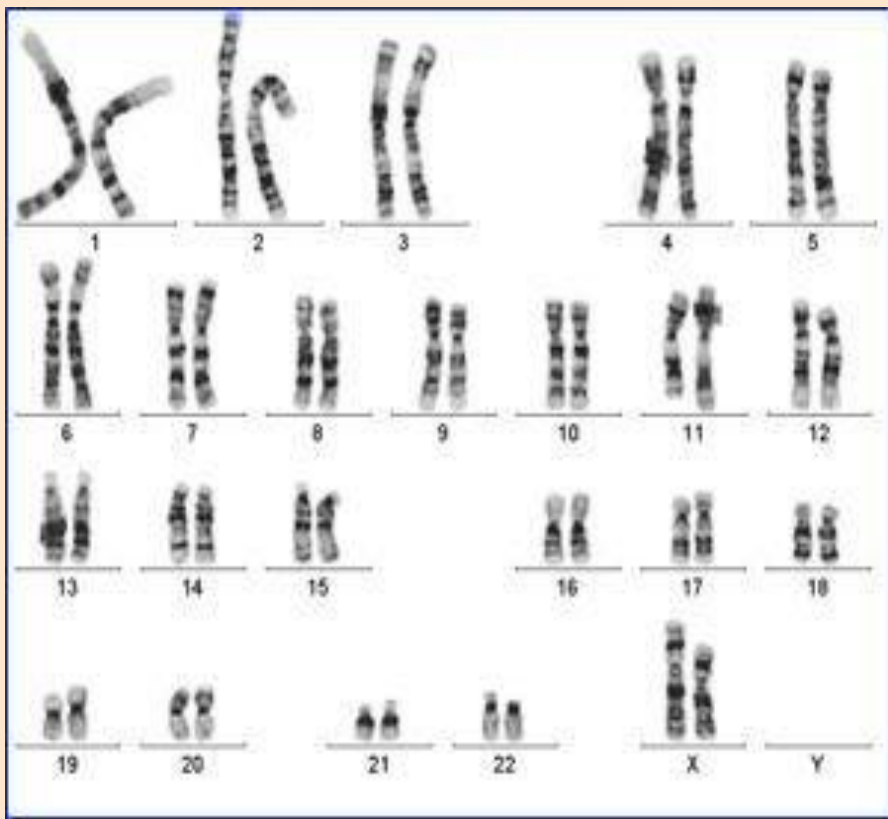


FACEPE

Fundação de Amparo à Ciência e
Tecnologia do Estado de Pernambuco

**Um pouco sobre
o problema....**





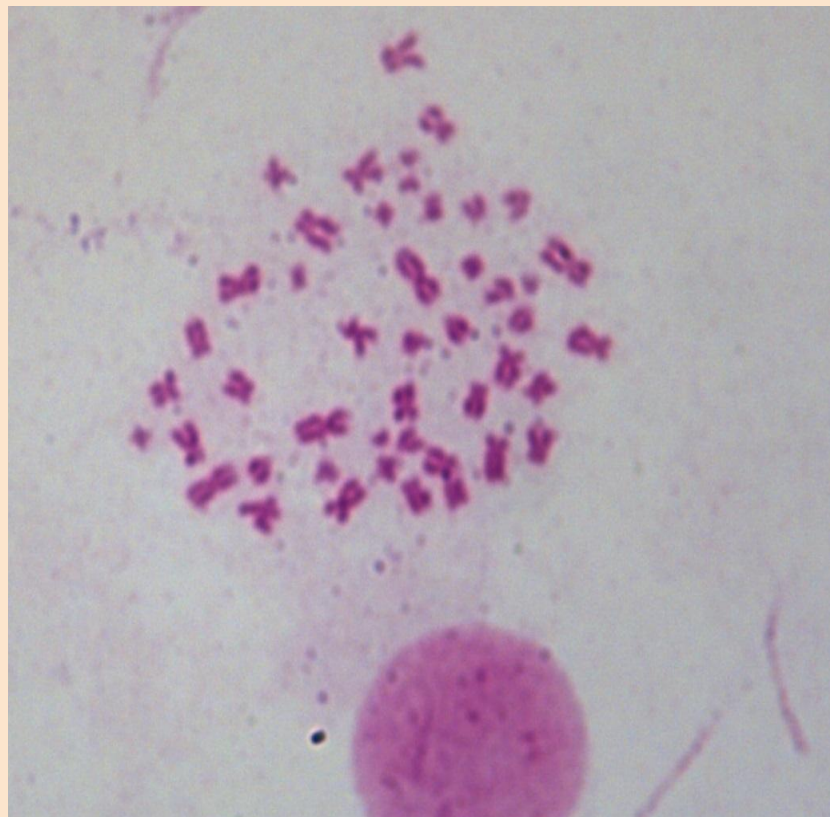
CRCN|NE

CENTRO REGIONAL DE CIÊNCIAS
NUCLEARES DO NORDESTE

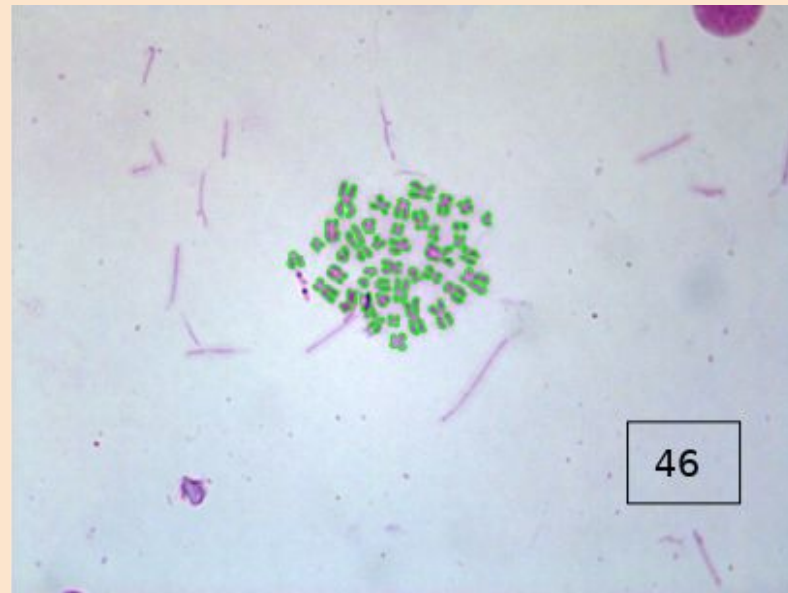


CRCN|NE

CENTRO REGIONAL DE CIÊNCIAS
NUCLEARES DO NORDESTE

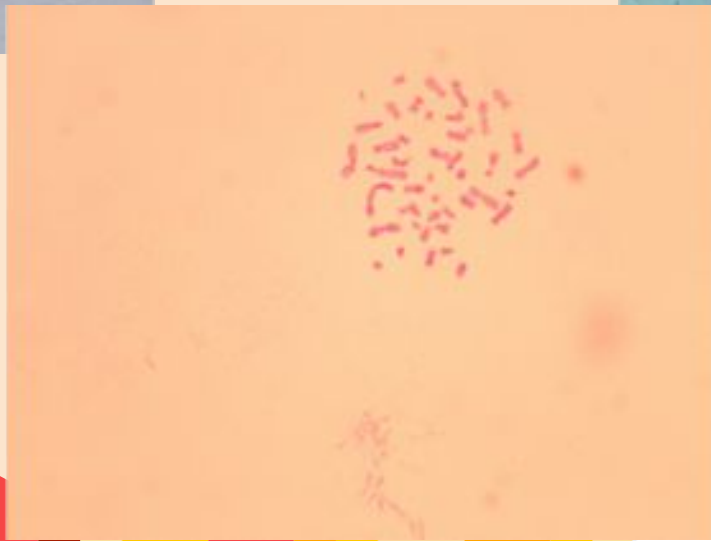
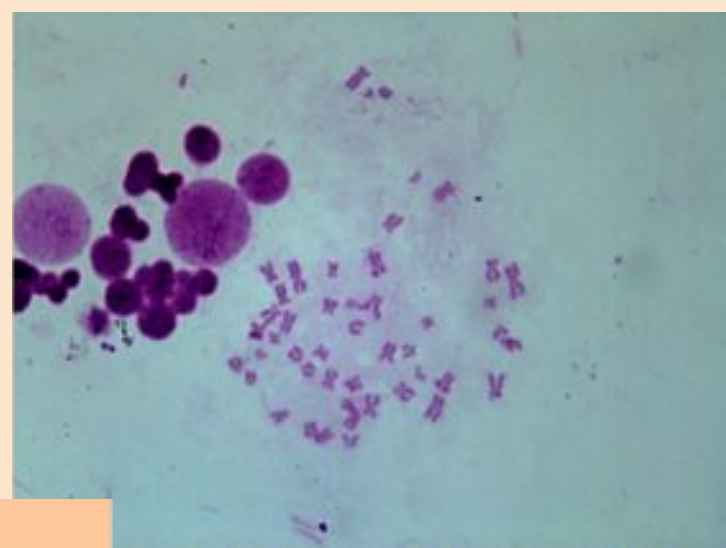
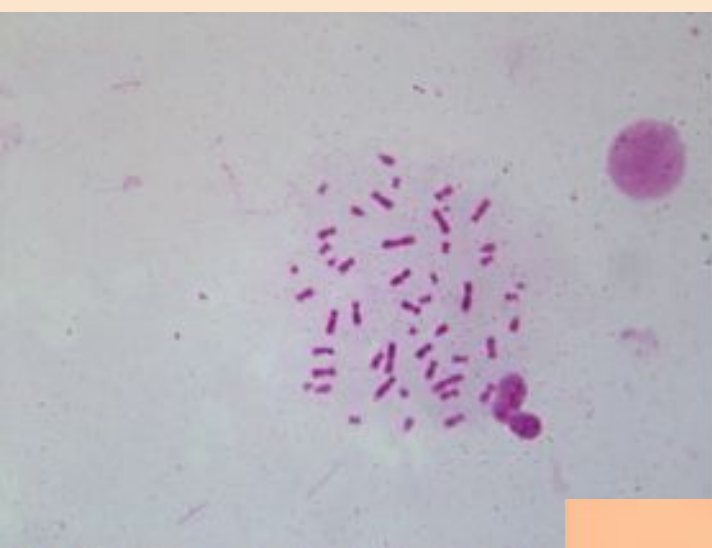


Solução esperada



Problemas encontrados







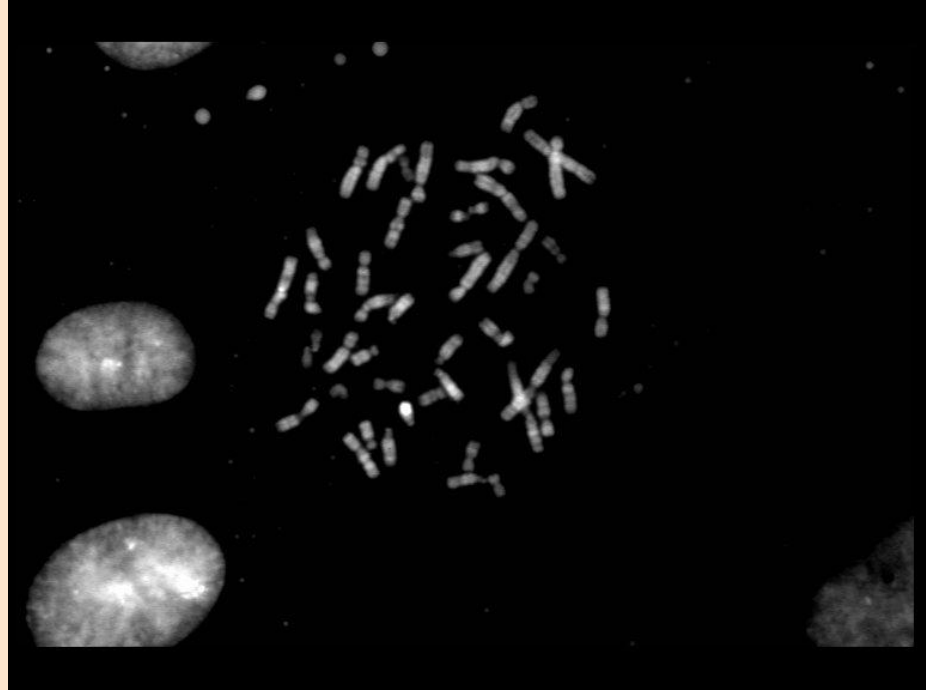
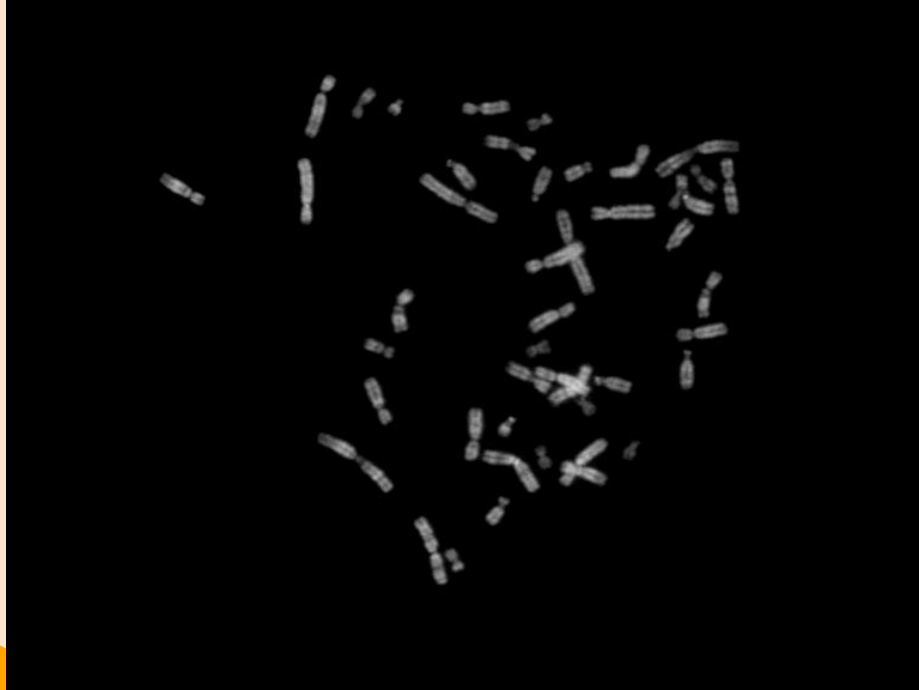
“

*Antes de tudo, precisamos
fazer um levantamento do
que vem sendo feito no
estado da arte.*

**BiolmLab o DataSet
mais utilizado nesse
problema**

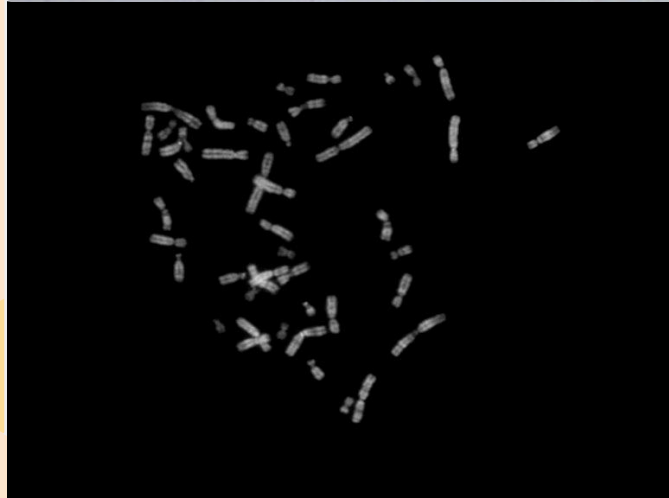
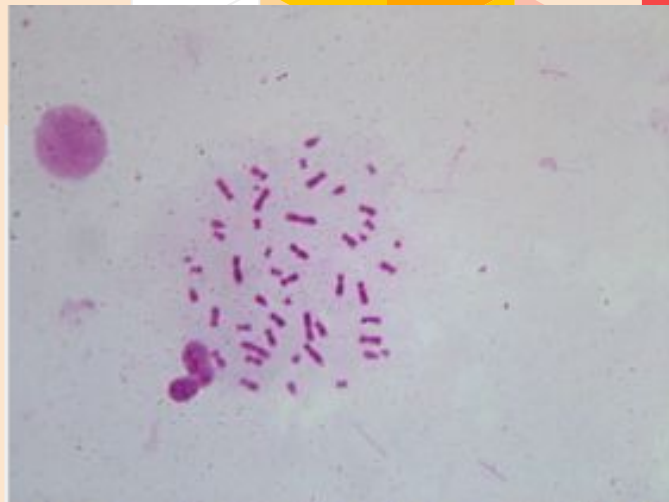


Laboratory of Biomedical Imaging (BioImLab)



CRCN vs BioImLab

- ◀ 74 imagens
 - ◀ Sem rotulação
- ◀ Método de Coloração
- ◀ 114 imagens
 - ◀ Com rotulação
- ◀ Fluorescencia com o bandeamento Q.

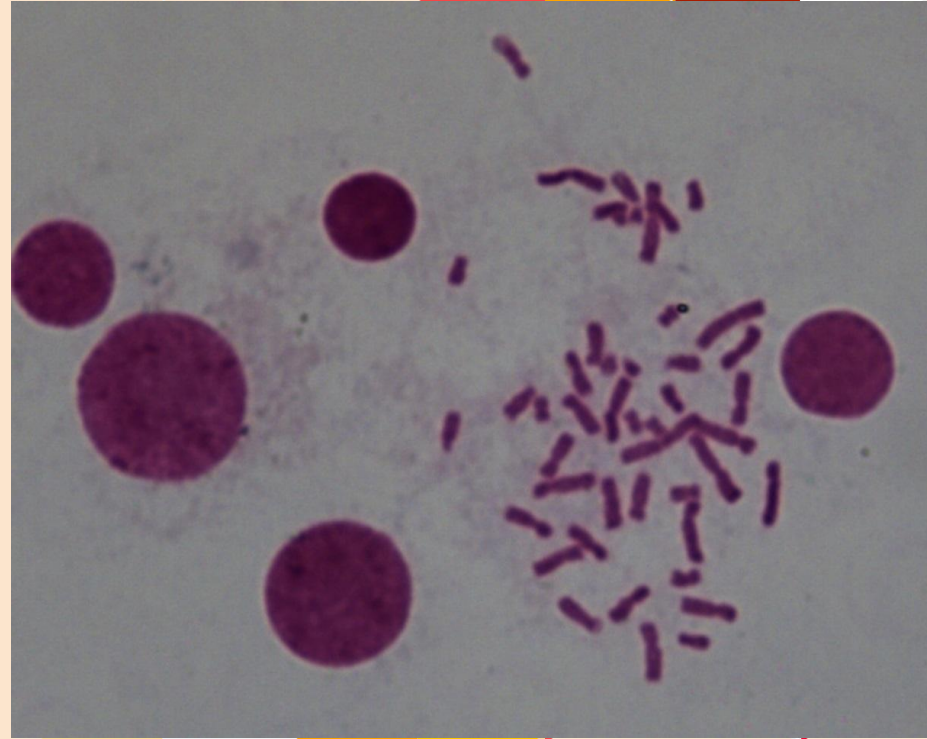


Base do CRCN-NE



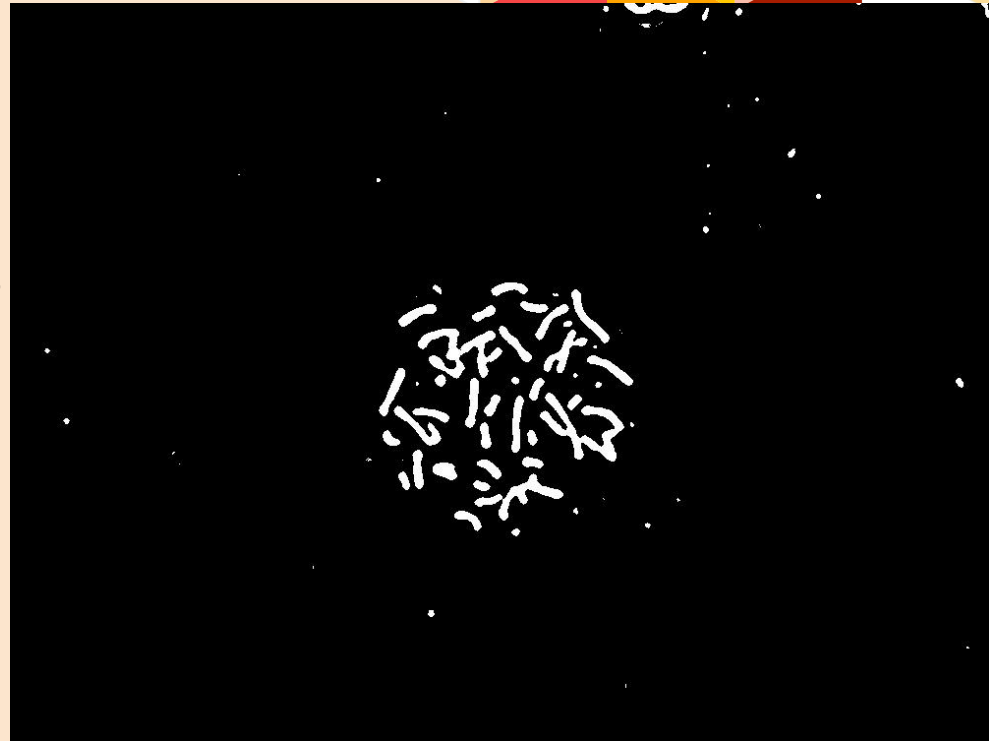
Base do CRCN-NE

- ▶ 74 imagens de metáfases, adquiridas no Centro Regional de Ciências Nucleares do Nordeste (CRCN-NE)
- ▶ Rotulações realizadas por especialistas



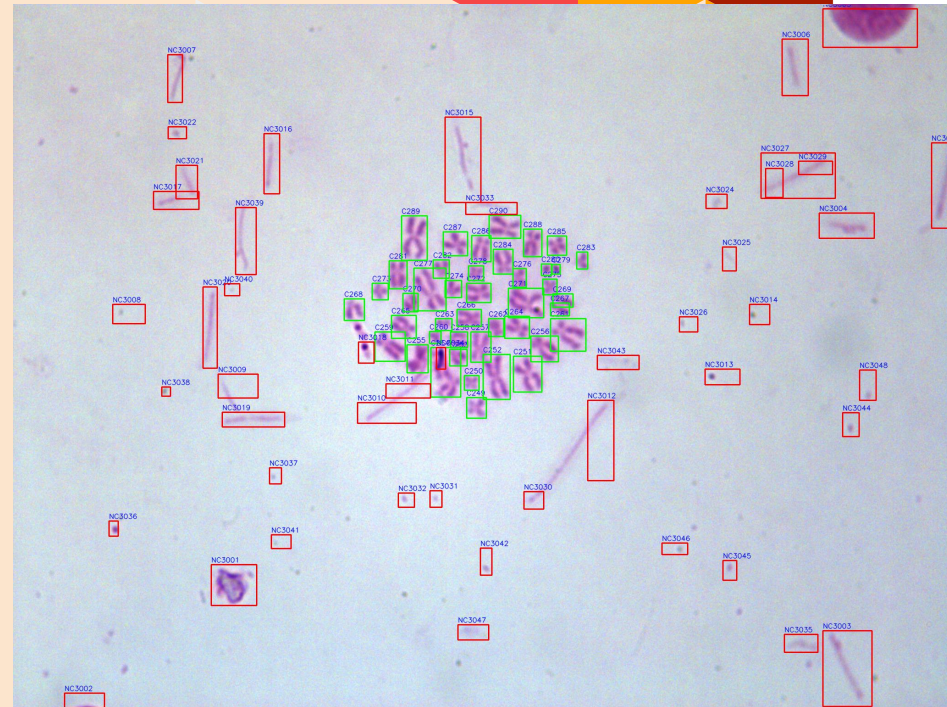
Imagens Segmentadas

- ◀ Imagens segmentadas, resultantes do método adaptativo
- ◀ Outros métodos foram testados



Regiões de interesse

- ▶ Marcação das regiões de Cromossomo e Não-Cromossomo com a ferramenta LabelMe;
- ▶ Captura de características em cada uma dessas regiões.



Base CRCN-NE

zenodo

Search



Upload

Communities

Log in

Sign up

May 26, 2019

Dataset

Open Access

CRCN-NE Chromosomes DataSet

Andrade, M. F. S.; Cordeiro, F. R.; Silva, J. J. G.; Lima, F. F.; Hwang, S.; Macário, V.

This dataset was obtained in partnership with The Biological Dosimetry Laboratory(CRCN-NE) and The Computing Department (DC) of the Rural Federal University of Pernambuco (UFRPE), both located in the city of Recife - Brazil.

Here are 74 images of colored metaphases acquired with a Leica microscope. The labeling of these metaphases will also be found, in this case, their expected results. In addition, we used the adaptive thresholding algorithm on this dataset, the result is also found in this file. So, in total are 74 images of colored metaphases and 74 images of segmented metaphases.

Finally, cuts of regions of chromosomes and non-chromosomes will be found for both the color image and the segmented image. Being present in the same quantity for both distributions, since they were obtained by the same technique.

- Cuttings of chromosome regions occurred using the Python OpenCv library, which swept the labeled image and captured the regions concerning the chromosomes. With these coordinates it was possible to perform the mapping for the colored and segmented images, making the extraction of these areas of interest.
- The cuts of the regions of non-chromosomes occurred with the aid of the LabelMe tool, in it was carried out the marking of 46 random points in the image. With this tool, it is possible to obtain an xml with the coordinates of these markings, which were used to extract the areas of interest.

Files (1.9 GB)

56

views

47

downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

May 26, 2019

DOI:

DOI [10.5281/zenodo.3229434](https://doi.org/10.5281/zenodo.3229434)

Keyword(s):

chromosomes

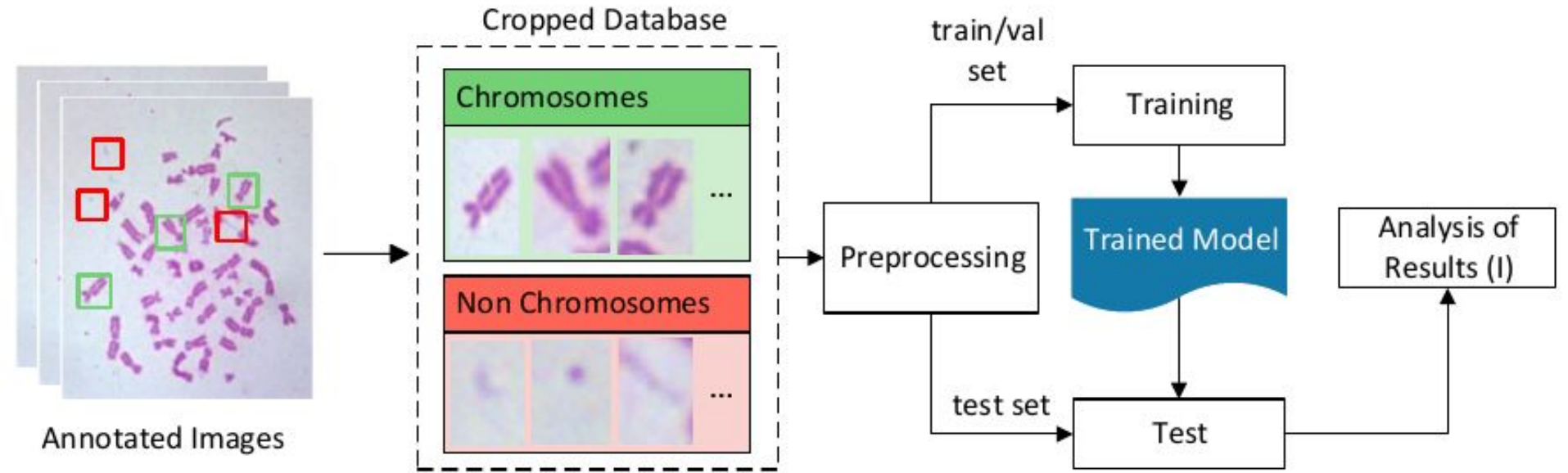
crn

metaphases

Metodologia

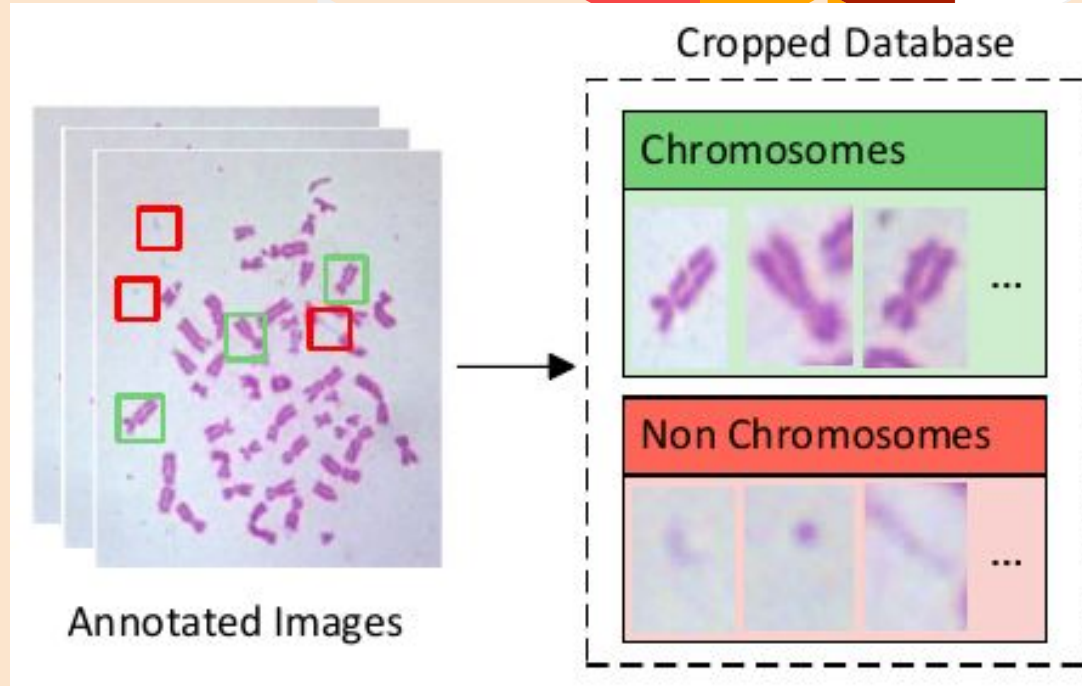


Metodologia



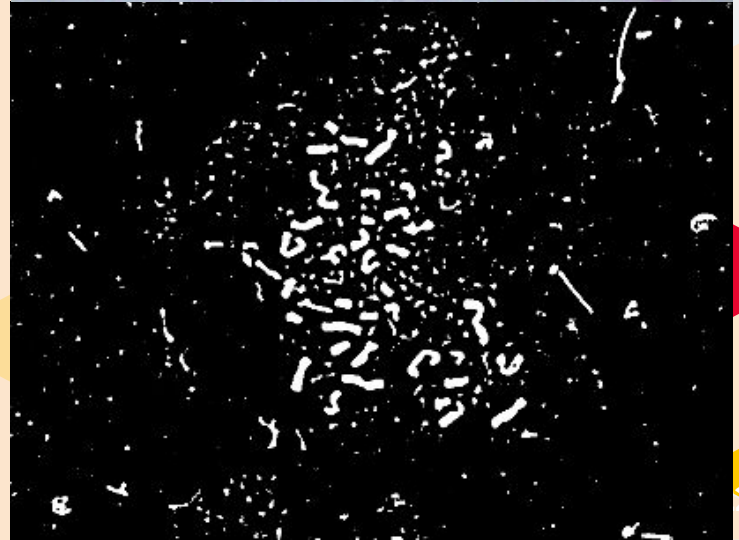
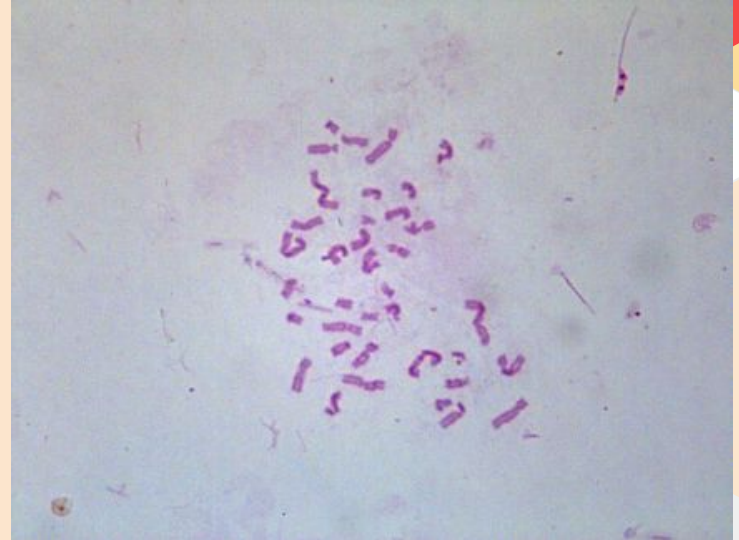
Entradas

- 60% (48 imagens) para treinamento, 20% (15 imagens) para teste e 20% (11 imagens) para validação;

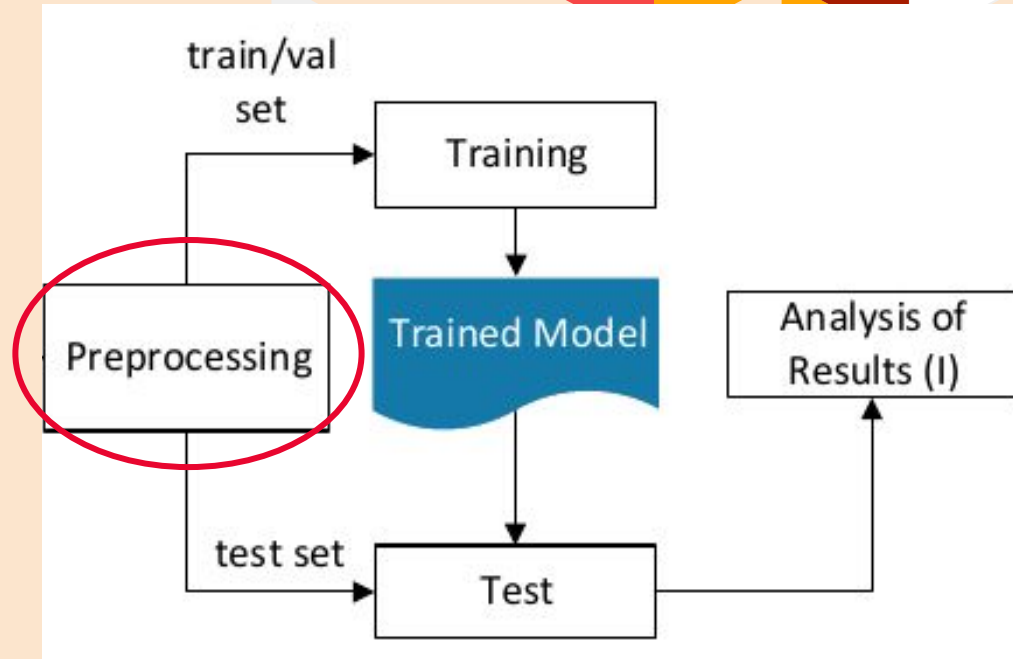


Entradas

- ▶ Imagens coloridas ou segmentadas;
- ▶ Recortes sendo do tipo cromossomos ou não cromossomos.

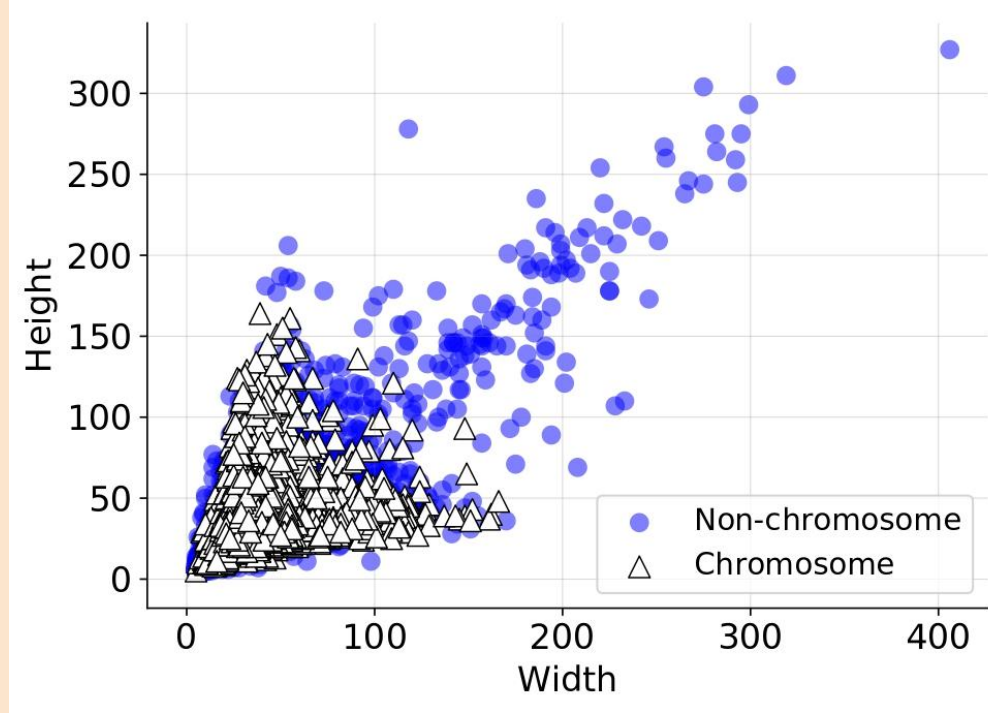


Pré-processamento

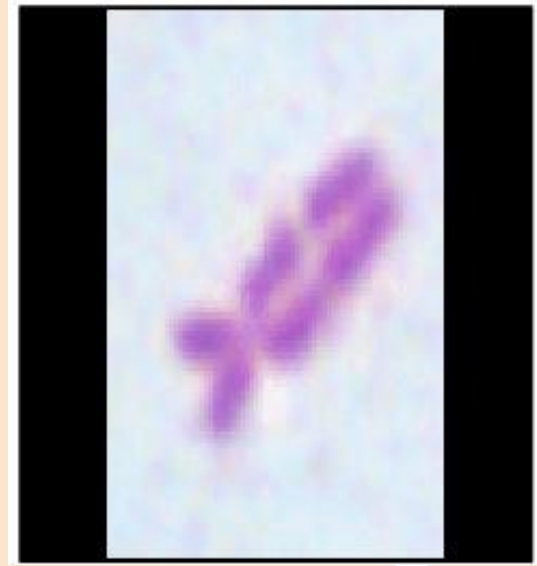
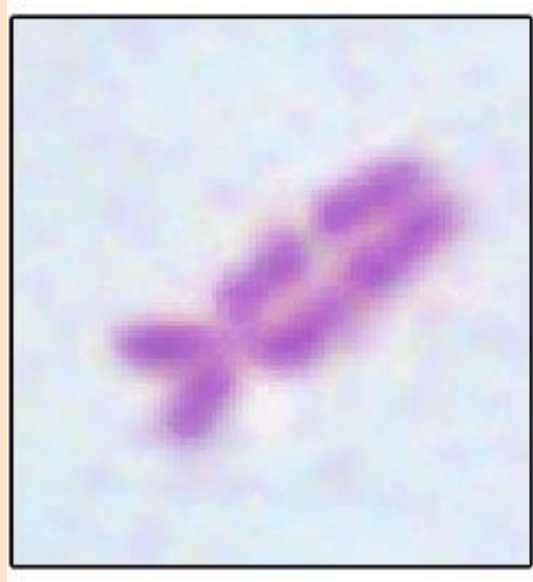
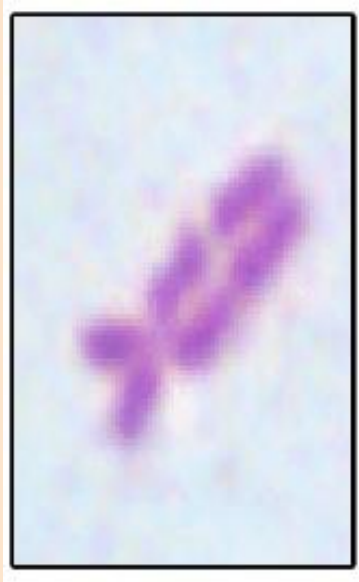


Pré-processamento

- ▶ Recortes com tamanhos variados;
- ▶ Entrada das CNN precisam ter um tamanho mínimo e pré-definido;

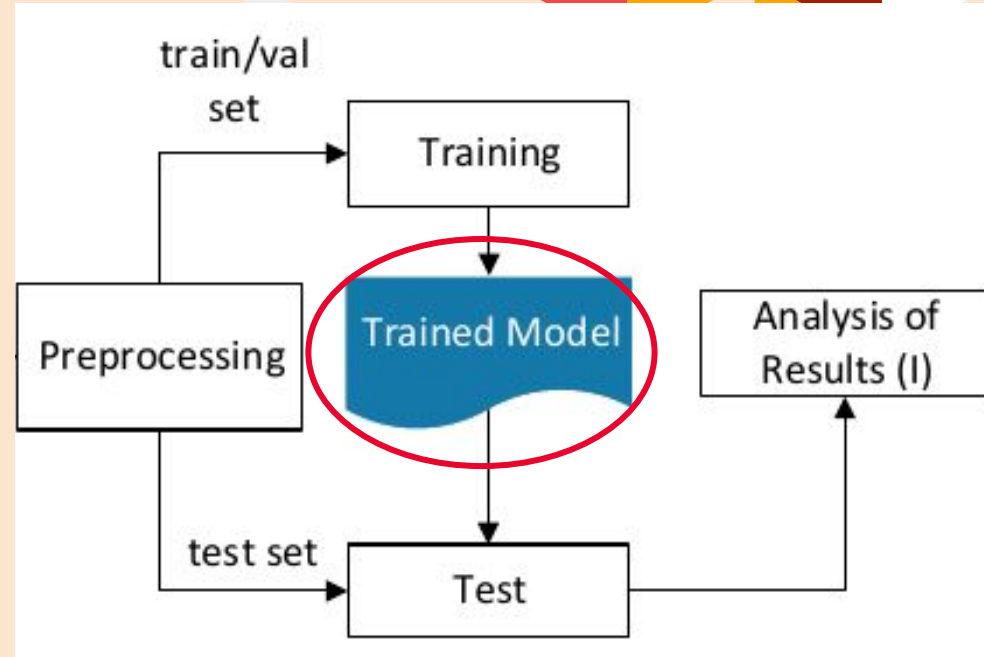


Pré-procassamento



Modelos treinados

- ▶ VGG16
- ▶ VGG19
- ▶ Xception
- ▶ MobileNet
- ▶ Inception_v3
- ▶ Sharma (artigo das interwebs)



**O que vamos
fazer com todos
esses modelos?**





Quais os principais problemas?

- ◀ Uma base relativamente pequena para utilizar CNN;
- ◀ Elevado tempo exigido para o treinamento;
- ◀ Tentativa falha de desenvolvimento de modelos;
 - ◀ Poucos modelos criados (e descritos detalhadamente) para o problema de cromossomos.

Transfer Learning

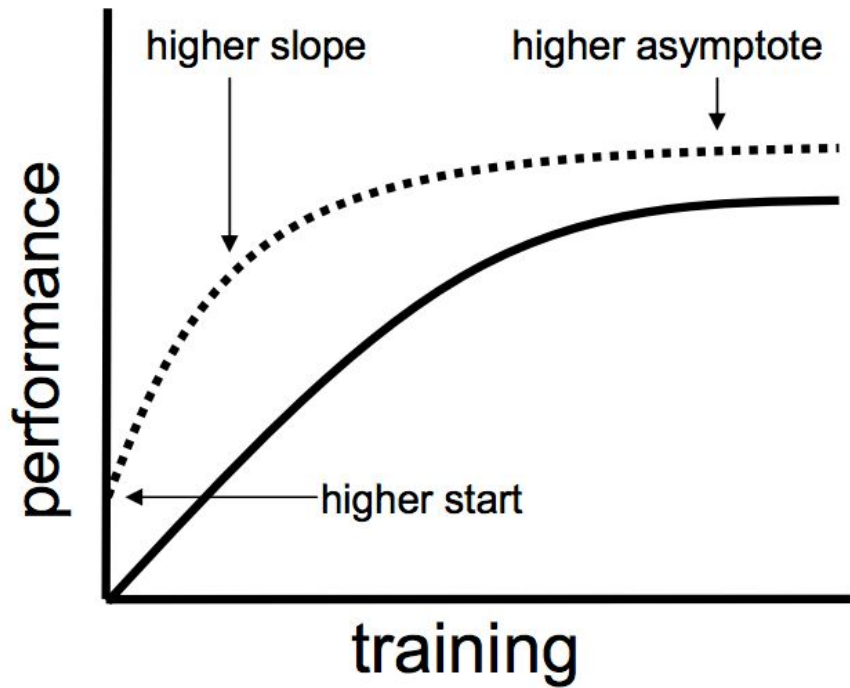




“

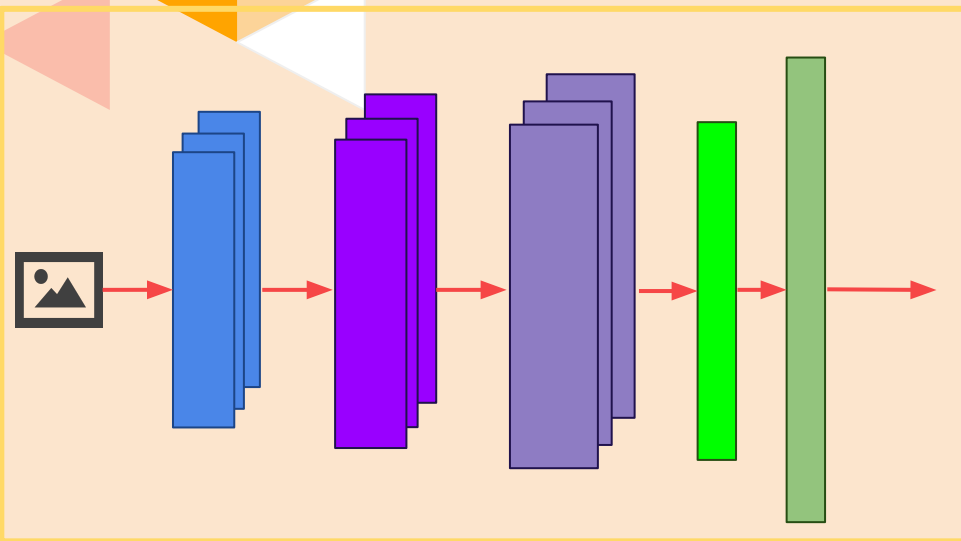
Transfer learning and domain adaptation refer to the situation where what has been learned in one setting ... is exploited to improve generalization in another setting.

Page 526, Deep Learning, 2016



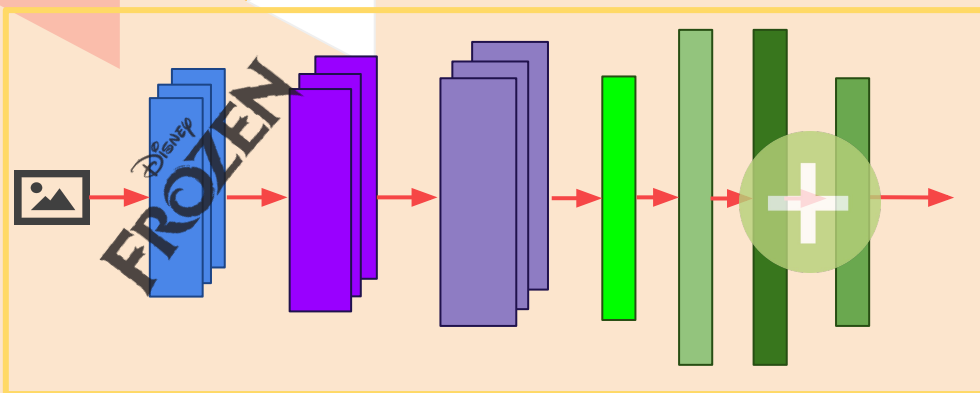
..... with transfer
— without transfer

Regrinhas para utilizar o TL



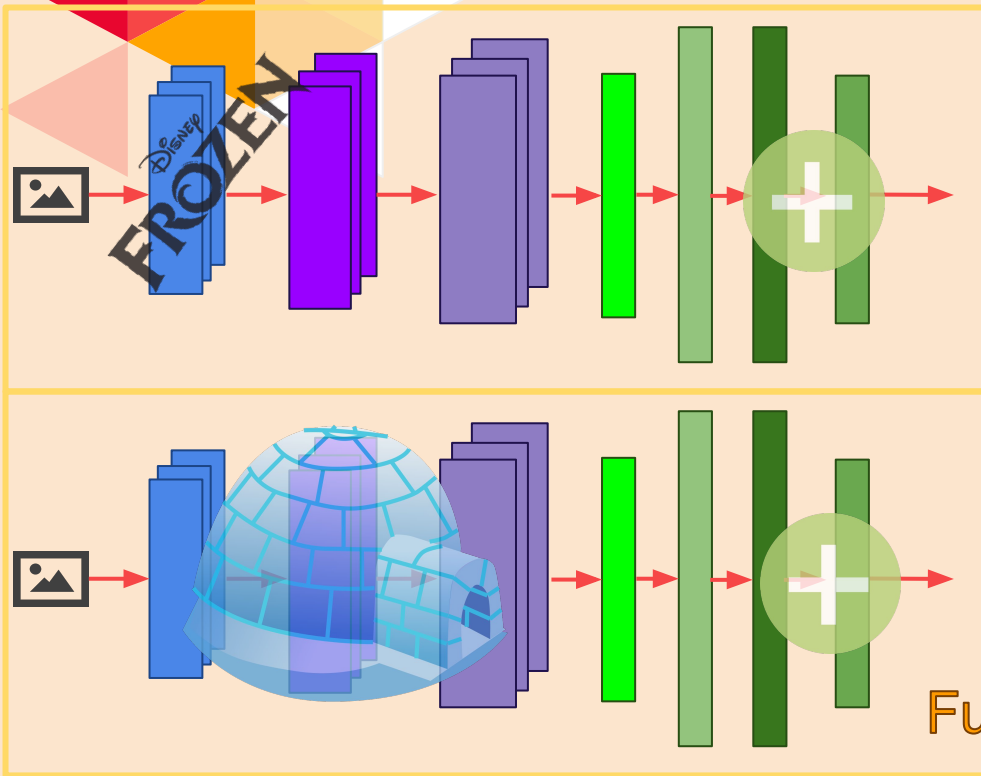
	Conjunto de dados pequeno e similar ao do modelo pré-treinado	Conjunto de dados grande e similar ao do modelo pré-treinado
Similaridade dos dados	Conjunto de dados pequeno e muito diferente do modelo pré-treinado	Conjunto de dados grande e muito diferente do modelo pré-treinado
	Quantidade de dados	


Regrinhas para utilizar o TL



	Conjunto de dados pequeno e similar ao do modelo pré-treinado	Conjunto de dados grande e similar ao do modelo pré-treinado
Similaridade dos dados	Conjunto de dados pequeno e muito diferente do modelo pré-treinado	Conjunto de dados grande e muito diferente do modelo pré-treinado
	Quantidade de dados	

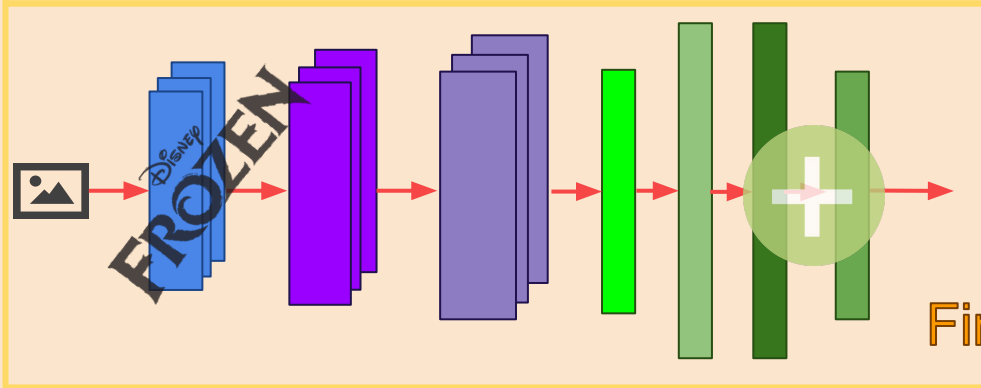
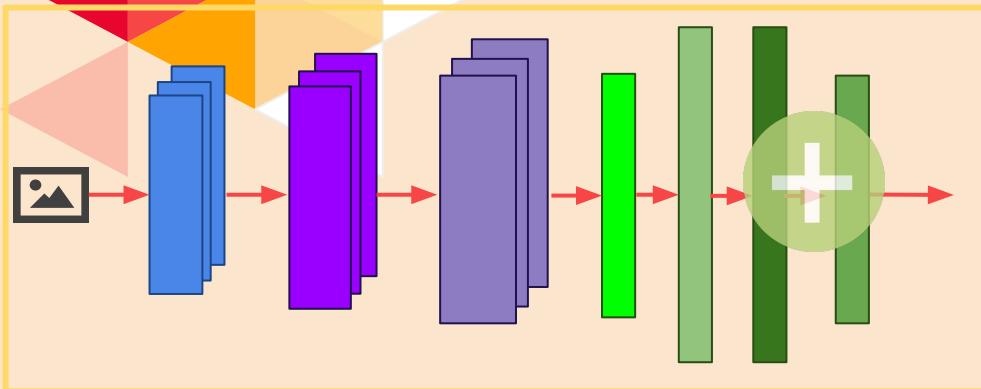
Regrinhas para utilizar o TL



Similaridade dos dados	 Conjunto de dados pequeno e similar ao do modelo pré-treinado	Conjunto de dados grande e similar ao do modelo pré-treinado
	Conjunto de dados pequeno e muito diferente do modelo pré-treinado	Conjunto de dados grande e muito diferente do modelo pré-treinado
Quantidade de dados		

Fully Connected (FC)

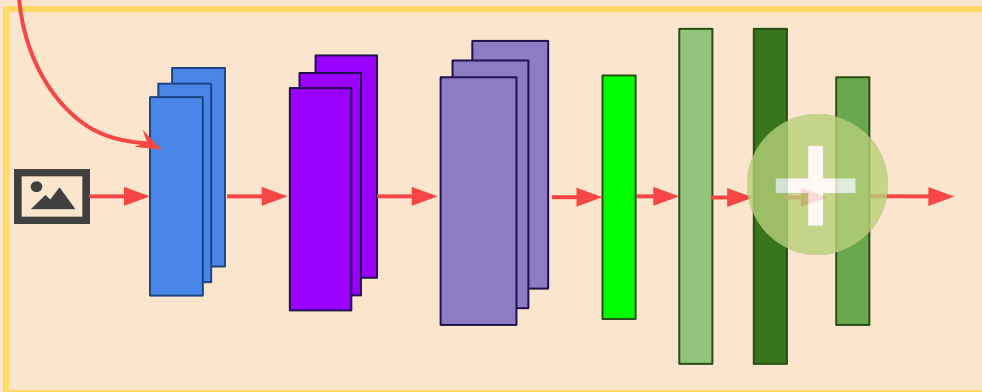
Regrinhas para utilizar o TL



Fine Tune (FT)

	Conjunto de dados pequeno e similar ao do modelo pré-treinado		Conjunto de dados grande e similar ao do modelo pré-treinado
Similaridade dos dados			
	Conjunto de dados pequeno e muito diferente do modelo pré-treinado		Conjunto de dados grande e muito diferente do modelo pré-treinado
		Quantidade de dados	

Regrinhas para utilizar o TL



Similaridade dos dados

Conjunto de dados pequeno e similar ao do modelo pré-treinado	Conjunto de dados grande e similar ao do modelo pré-treinado
Conjunto de dados pequeno e muito diferente do modelo pré-treinado	Conjunto de dados grande e muito diferente do modelo pré-treinado

Quantidade de dados

Os modelos utilizados

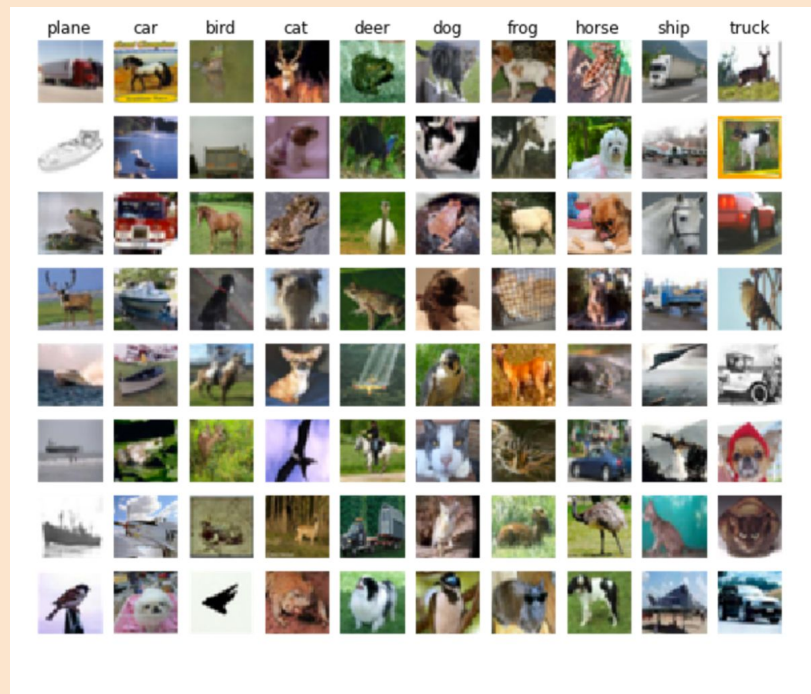
- ◀ Treinados na ImageNet
 - ◀ 14 milhões de imagens e 20000 classes

VGG

Inception

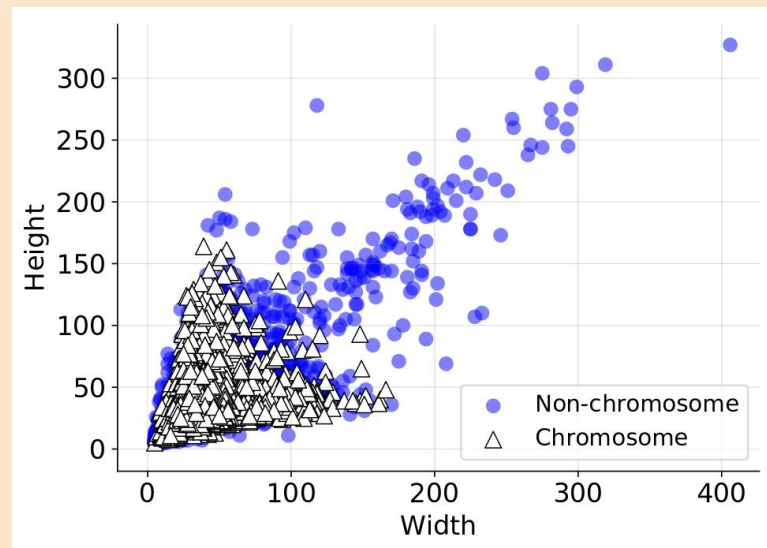
MobileNet

Xception



Lembra da etapa de normalização das entradas?

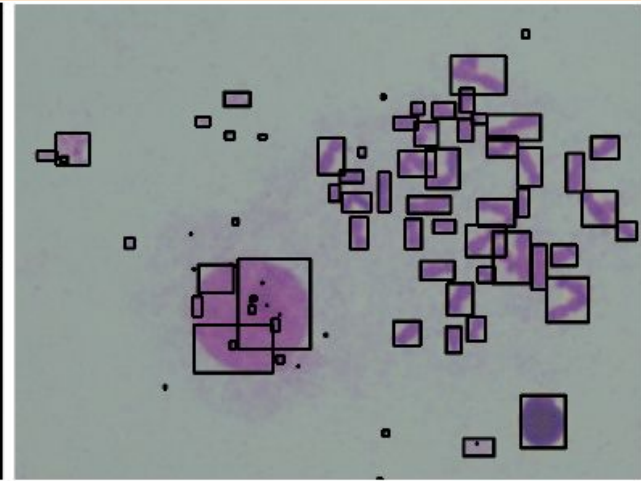
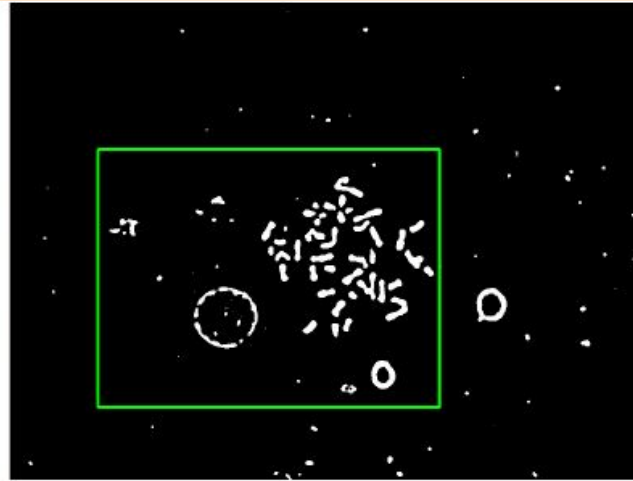
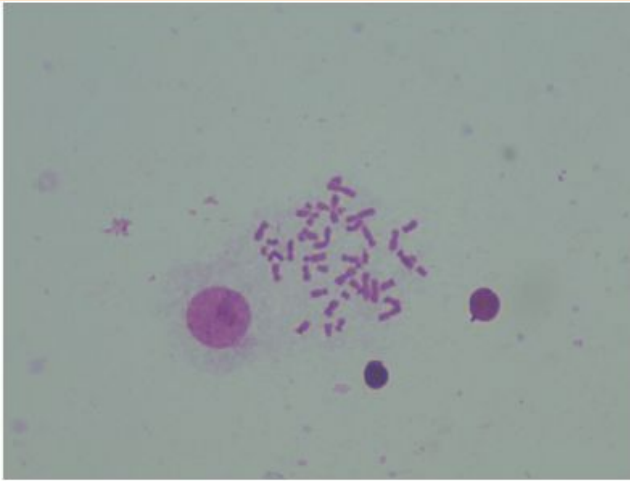
VGG	64x64
Inception	192x192
MobileNet	128x128
Xception	128x128



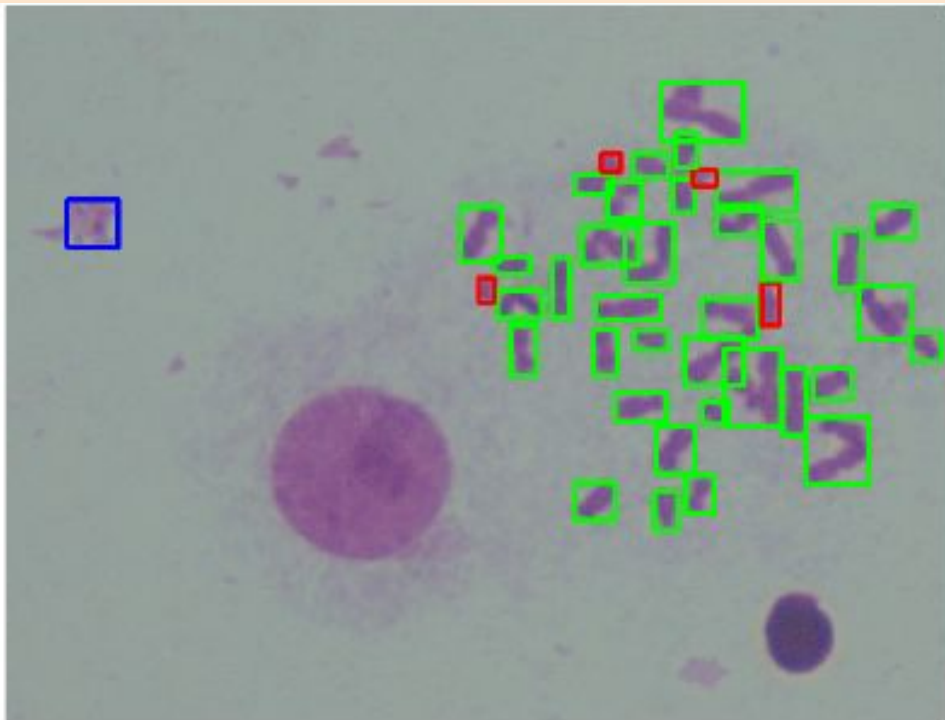
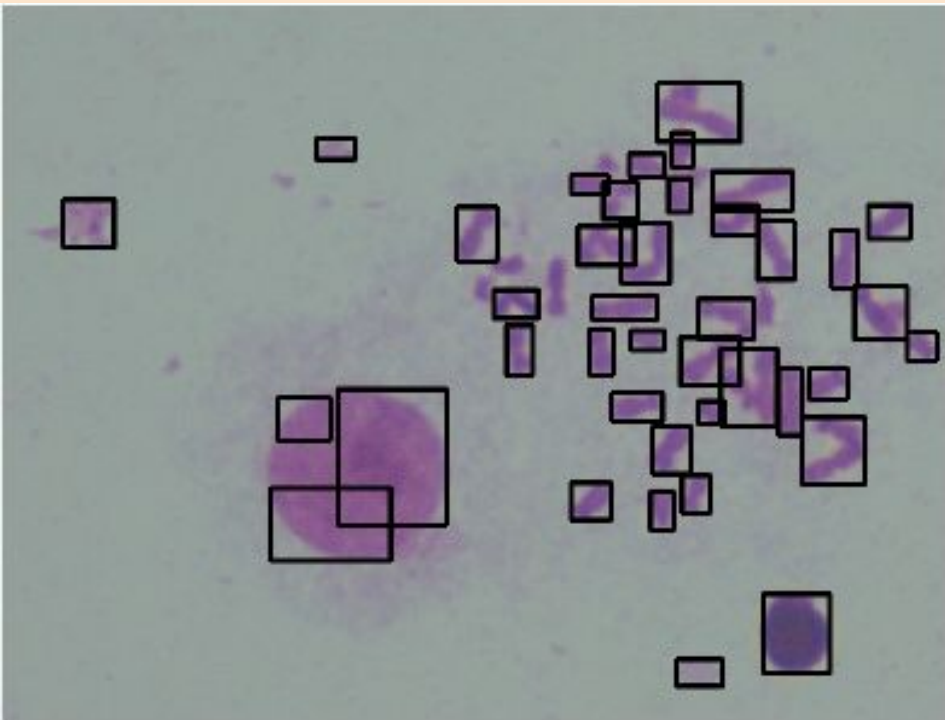
Resultados



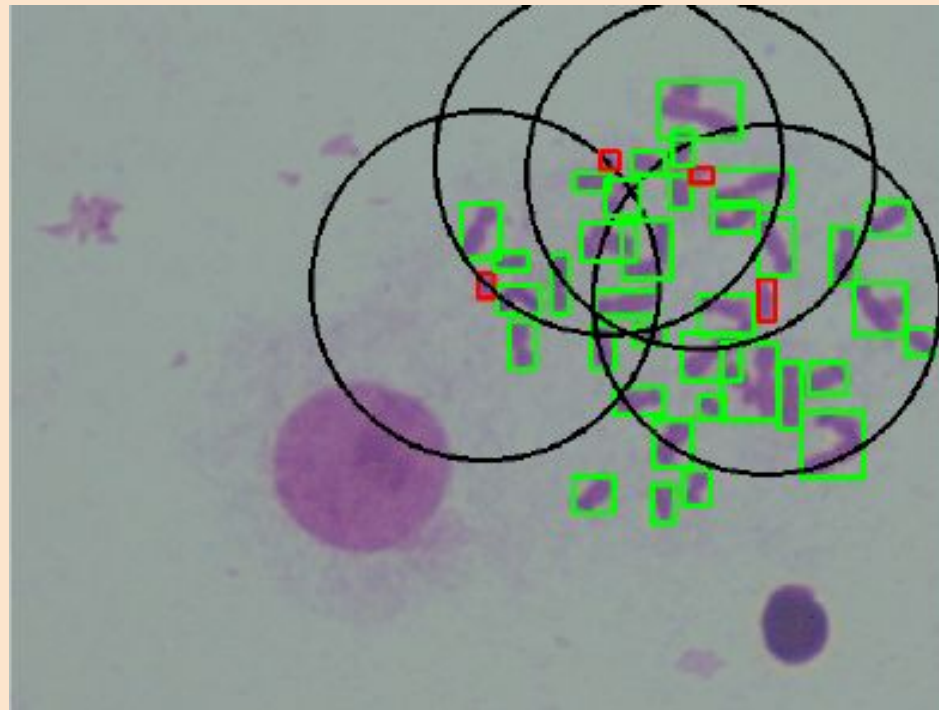
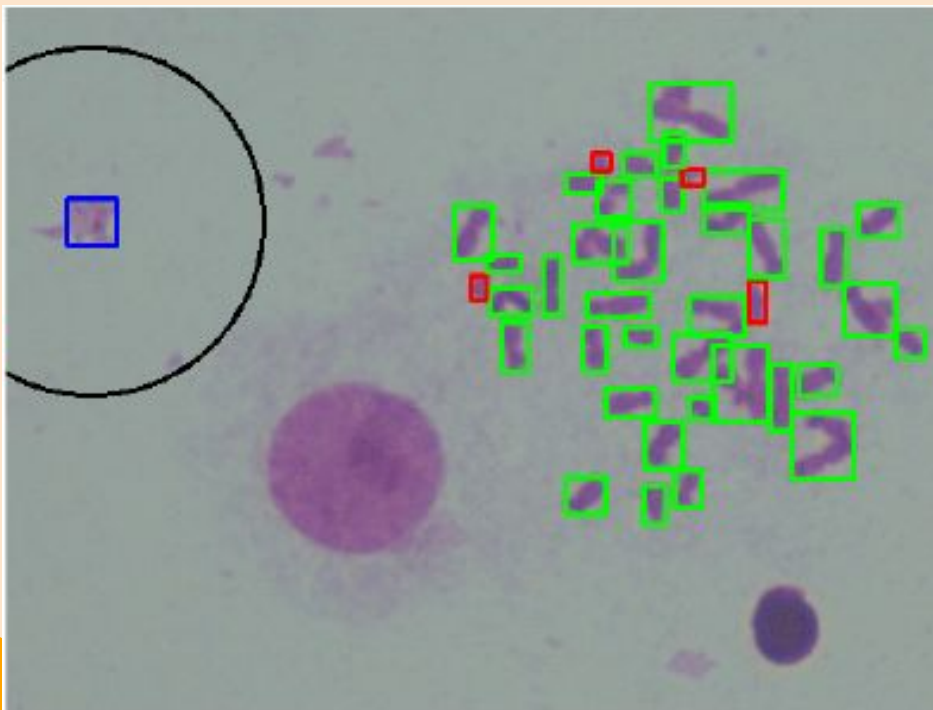
Resultados



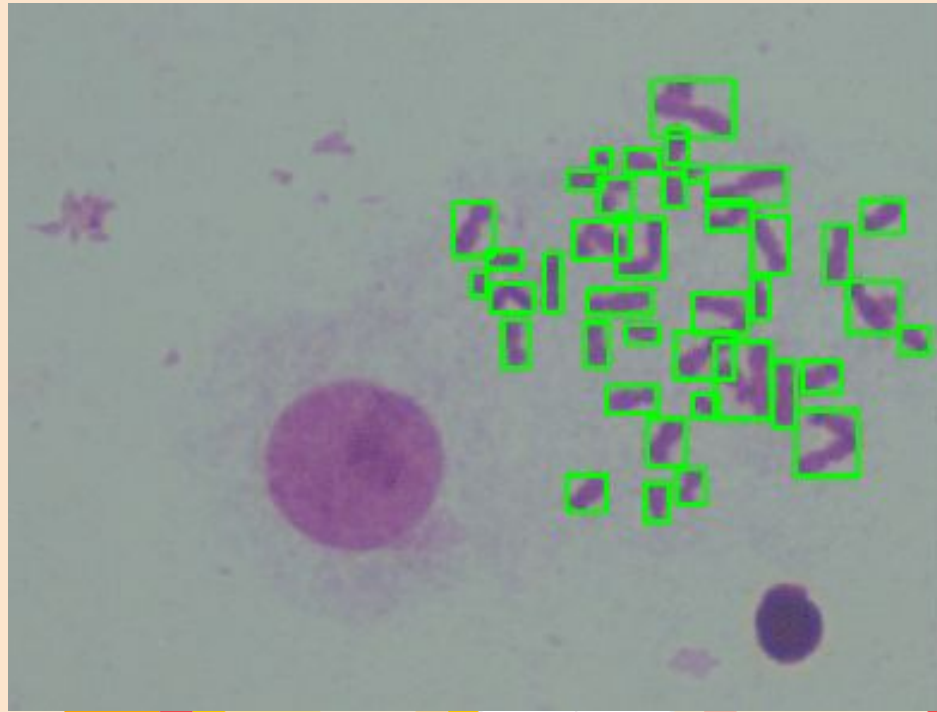
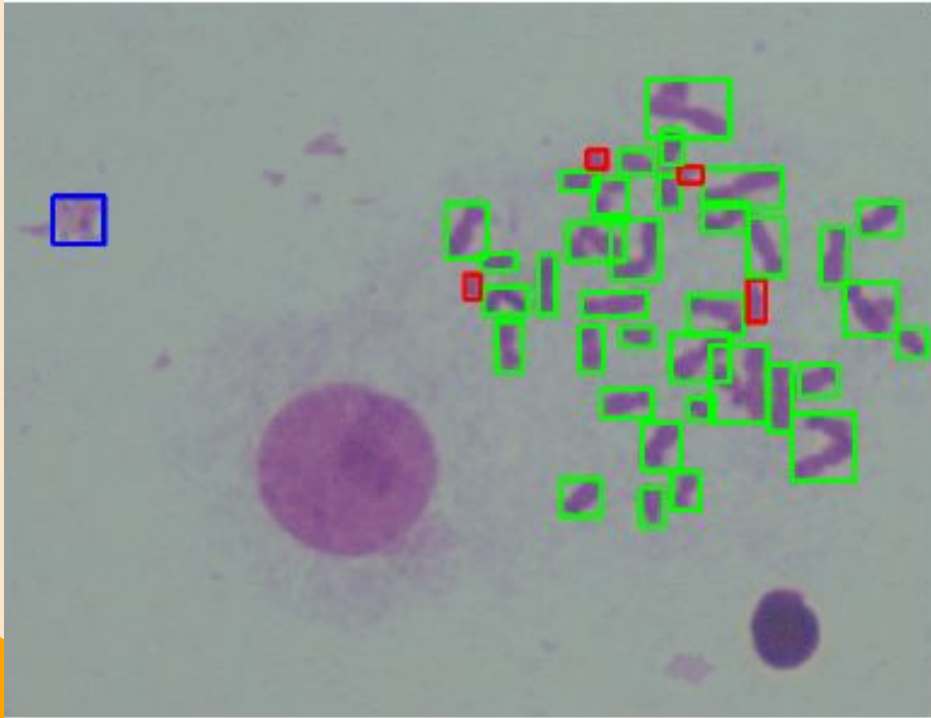
Resultados



Resultados



Resultados



Resultados

- ◀ Utilizar regiões segmentadas teve melhor desempenho;
- ◀ Arquiteturas baseadas em VGG tiveram resultados similares;

Architecture	Accuracy (%)	
	RGB	Binary
VGG16_FT	89.71±1.16	92.94±0.43
VGG16_TL	87.98±1.14	92.47±0.69
VGG19_FT	88.41±0.97	93.19±0.32
VGG19_TL	87.89±0.83	92.81±0.66
Inception_v3_FT	51.28±1.76	71.54±2.54
Inception_v3_TL	50.70±1.88	72.16±2.65
MobileNet_FT	52.21±5.61	80.39±2.14
MobileNet_TL	54.70±4.07	79.01±9.53
Xception_FT	53.02±4.16	77.87±9.15
Xception_TL	56.93±7.88	76.57±9.45
Sharma	59.50±19.34	66.11±18.39
MiniVGG	84.24±2.04	91.88±0.4

FT = congela apenas a primeira camada e treina as seguintes; TL = treina apenas as últimas camadas totalmente conectadas.

Conclusão





Conclusão

- ▶ Utilizar as imagens segmentadas, trouxeram melhores resultados, comparado com as imagens RGB;
- ▶ A qualidade da imagem segmentada impacta o resultado final;
- ▶ O tamanho utilizado para os recortes, afetam drasticamente a acurácia;
- ▶ O modelo VGG 16 com *transfer learning* do tipo FT, obteve os melhores resultados com 93,19% de acurácia.



Conclusão

Transfer Learning permitiu utilizar Deep Learning sem precisar de uma base com milhões de amostras!!!

Obrigada!

Dúvidas?

- ▶ @NaanAdr
- ▶ fsandrade25@gmail.com