

Fruits-360

Grupo Neurobook

Eduardo Armelin

Ivan Bravin

Marcio Pinheiro Neris

Tiago Deliberali Santos

1. Objetivo

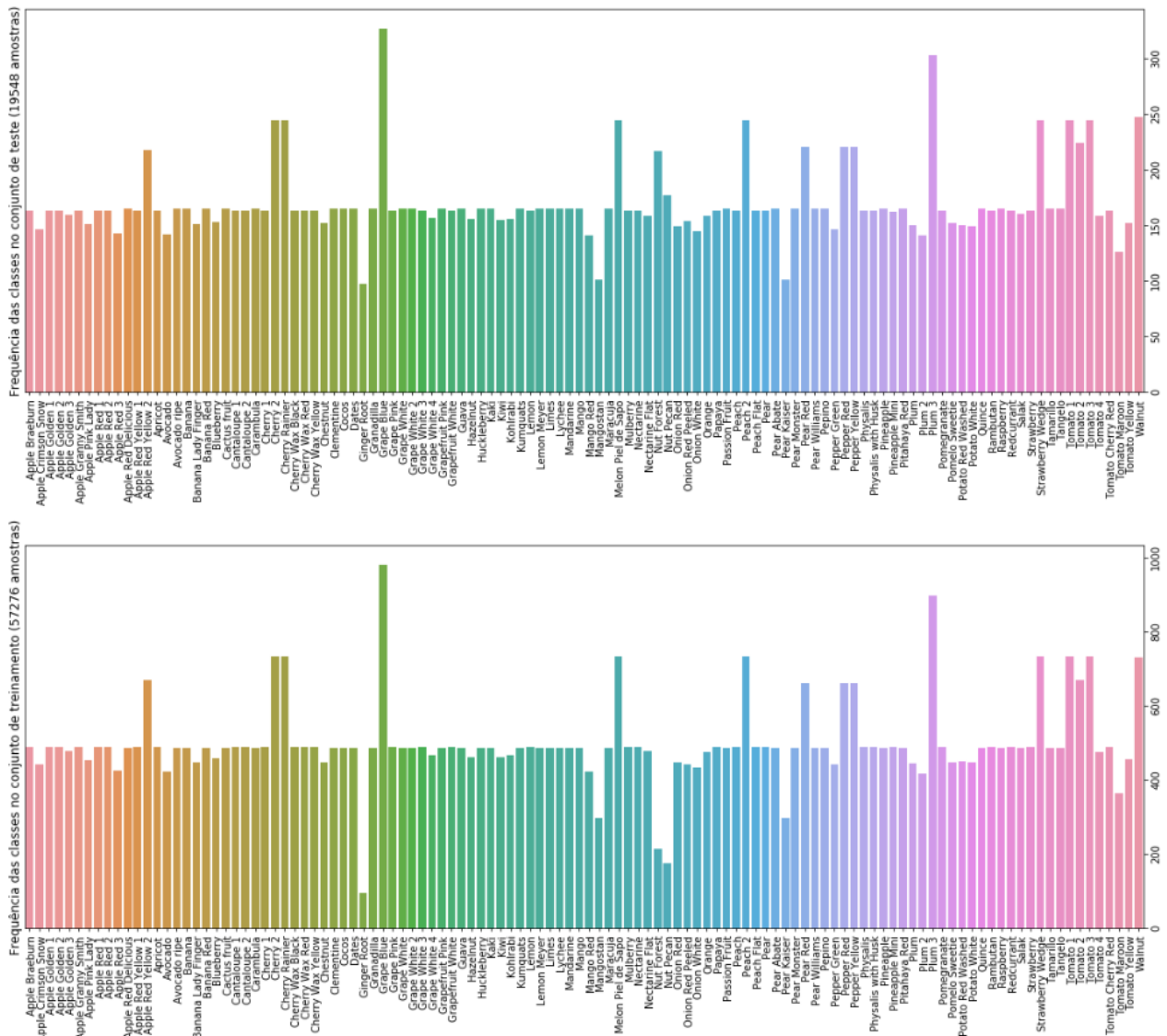
Classificação de fotos de diversas frutas. As fotos têm fundo branco, há apenas uma fruta por imagem e todas as imagens têm o mesmo tamanho, 100 pixels de altura por 100 pixels de largura. As frutas ocupam grande parte das fotos e não há escala. Ou seja, uma uva tem o mesmo tamanho que uma melancia.

2. Dados

São 114 tipos de frutas e legumes divididos em dois arquivos compactados, uma para teste e outra para treino, totalizando 76.824 imagens divididas da seguinte forma:

- 57.276 para treino
- 19.548 para teste

Os dados apresentam um leve desbalanceamento, conforme demonstrado abaixo.

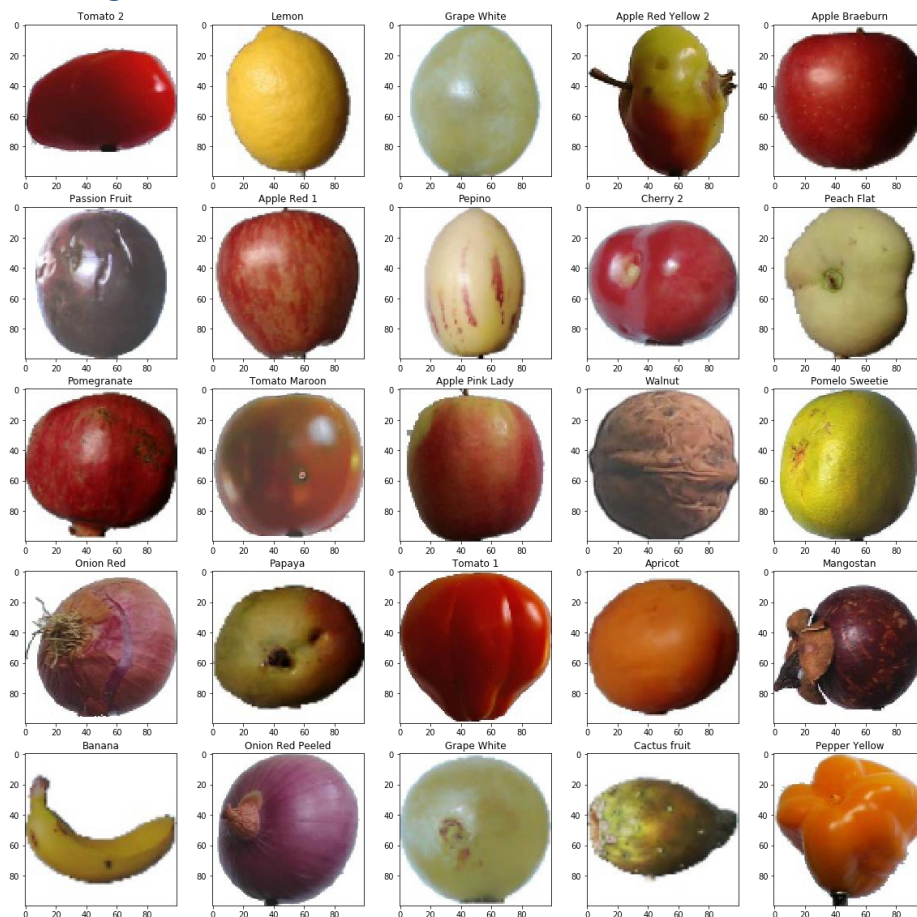


Uma vez que não foi disponibilizado dados para validação, os mesmos foram extraídos a partir dos dados de treino na proporção 80/20.

- 45.852 para treino
- 11.424 para validação
- 19.548 para teste

Os dados foram obtidos do Kaggle versão 2019.06.29.0.

2.1 Exemplos de imagem



3. Estratégia

Para a classificação das imagens propostas, foi utilizado duas abordagens, a construção de uma rede neural convolucional e *transfer learning* dos da *ImageNet*.

As redes utilizadas da *ImageNet* foram:

- InceptionV3
- MobileNet
- ResNet50
- VGG16

3.1 Transfer Learning

Para todas as redes utilizadas, foi utilizado os métodos de *Feature Extraction* e *Fine Tuning*. Para ambos os métodos o topo das redes foi removido e adicionado uma camada densa com ativação *softmax* adequada para a quantidade de classes do *dataset*.

3.2 Callbacks

Com o propósito de obter a melhor performance, foi utilizado em todas as redes *EarlyStopping* e *LearningRateScheduler*.

- *EarlyStopping*: encerra o treinamento quando a métrica monitorada para de melhorar com o passar das épocas
- *LearningRateScheduler*: ajusta o *learning rate* durante o treinamento a cada época

3.3 Parâmetros de treinamento

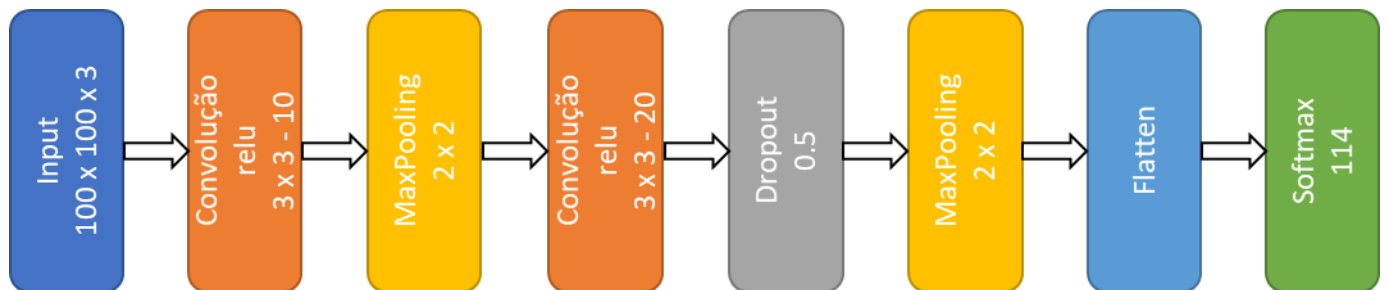
Todas as redes foram treinadas utilizando os seguintes parâmetros:

Optimizer	RMSprop
Initial Learning Rate	0.0002
Decay	1e-6
Batch Size	64
Epoch	50

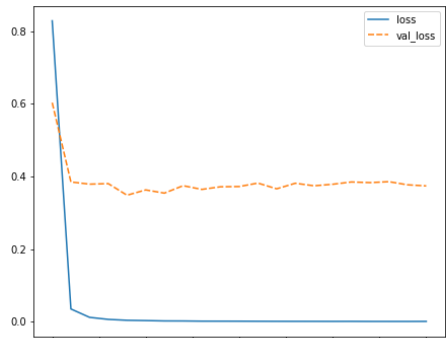
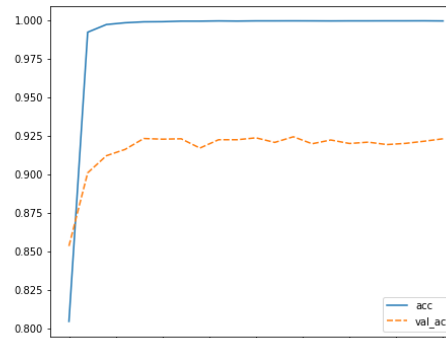
4. Treinando e avaliando os modelos na base de validação

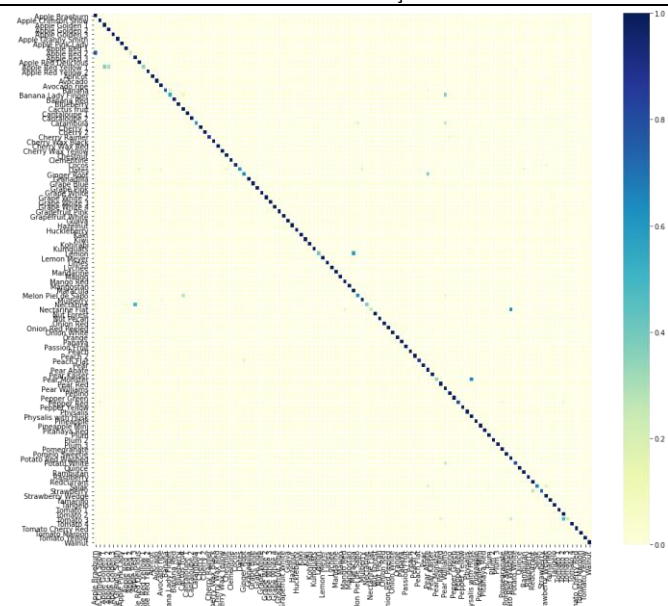
4.1 Homemade

A *Homemade* é uma rede neural convolucional com a seguinte arquitetura:



Abaixo é demonstrado o resultado ao submeter o conjunto de validação da rede treinada.

Parâmetros		Resultado de Treinamento	
Treináveis	Não Treináveis		
1.105.734	0		

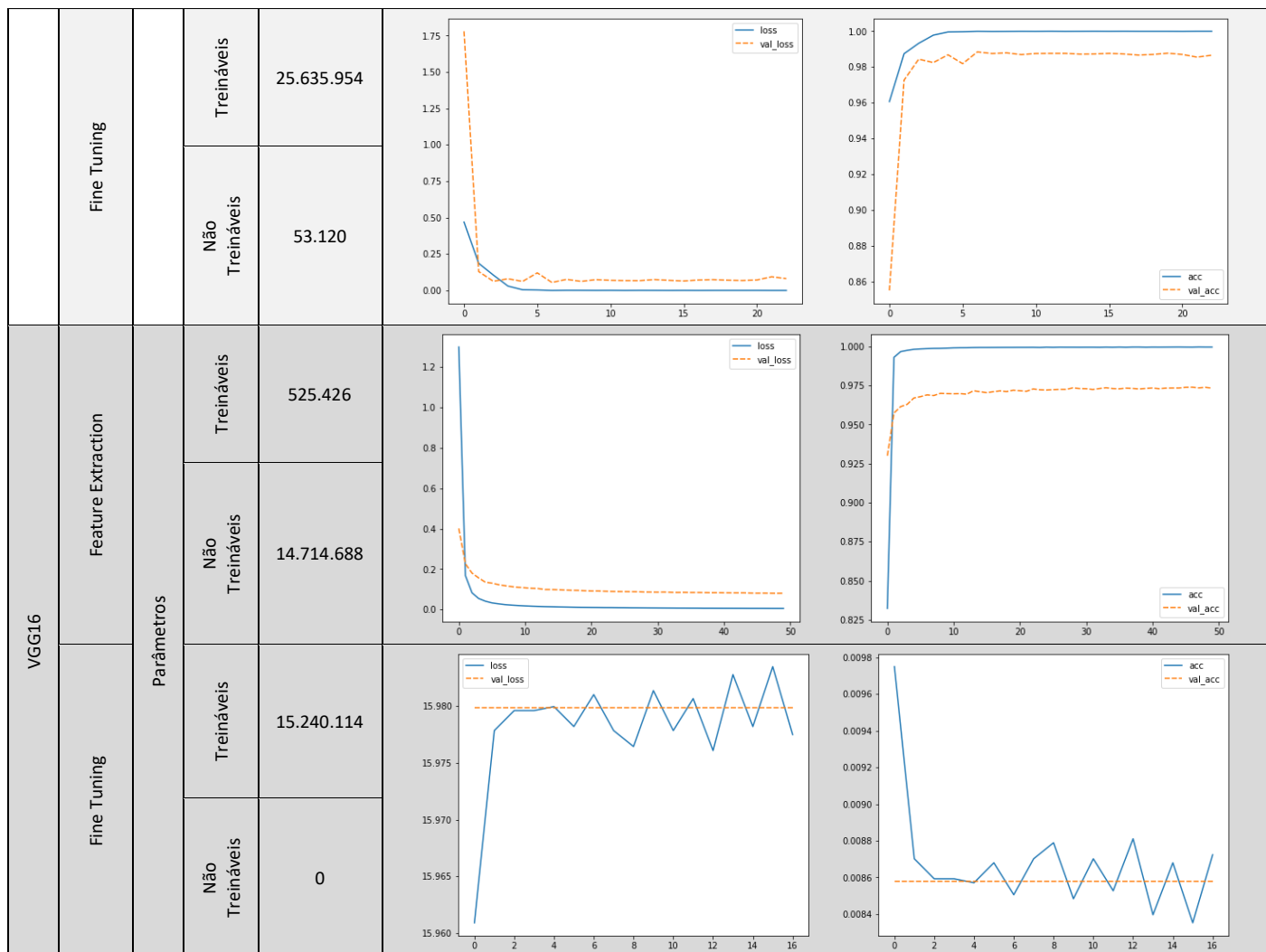
Acurácia		Matriz de Confusão Dados de Validação
Treinamento	Validação	
99,9542%	92,3319%	

4.2 Transfer Learning

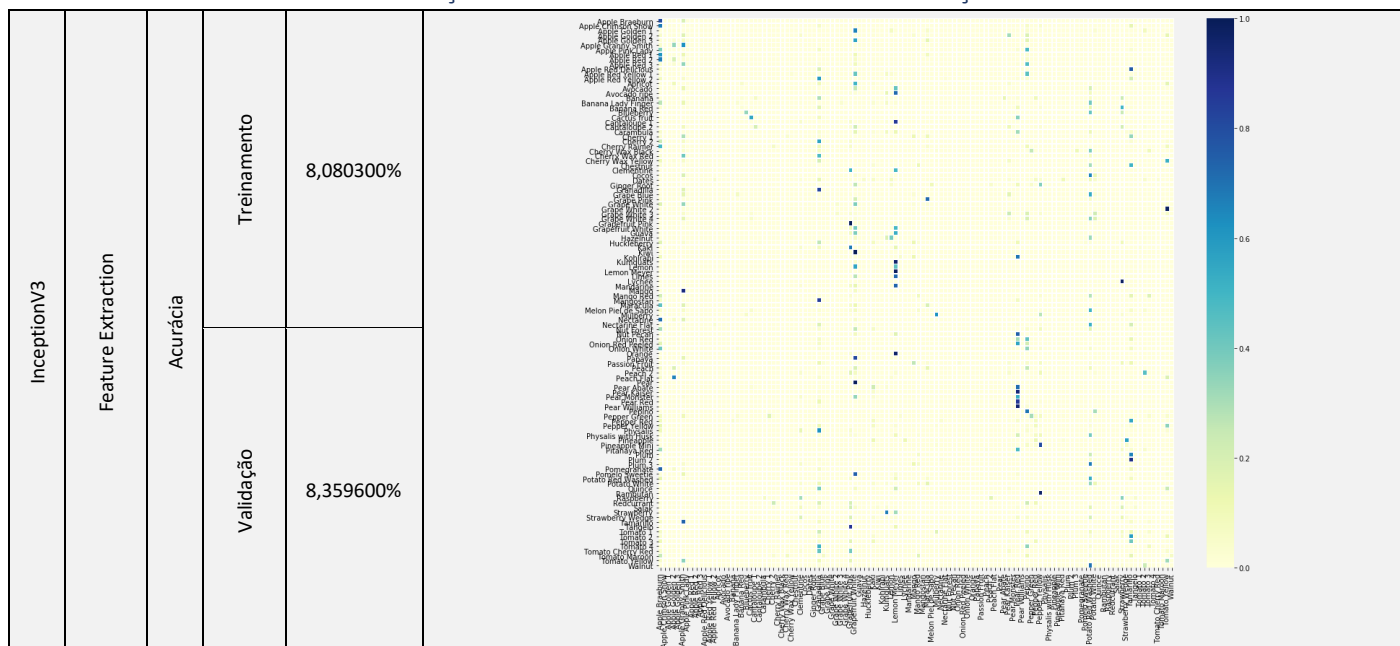
Para cada rede selecionada pré-treinada pela ImageNet, foi utilizado método de *Feature Extraction* e *Fine Tuning*. Abaixo é apresentado o resultado de todas as redes para cada método.

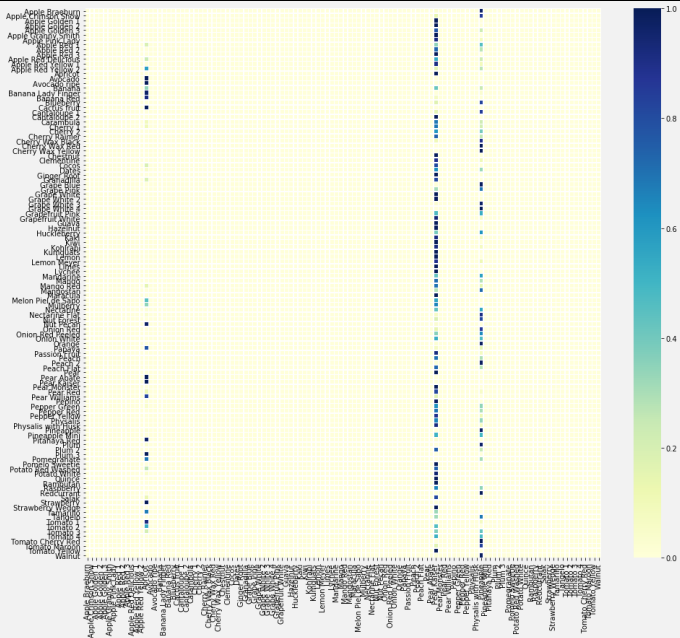
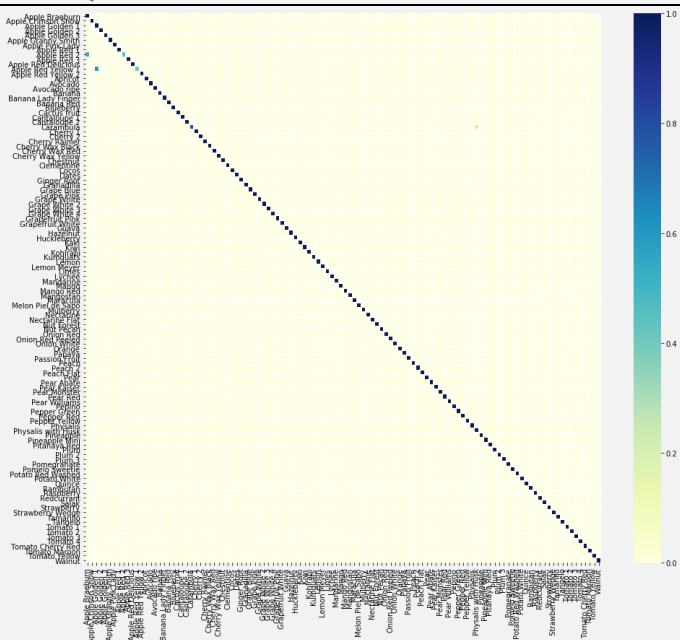
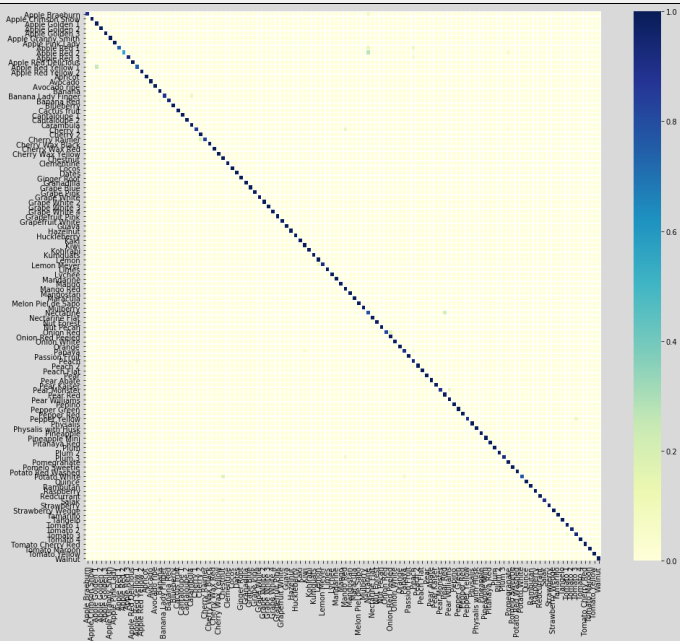
Resultado de Treinamento

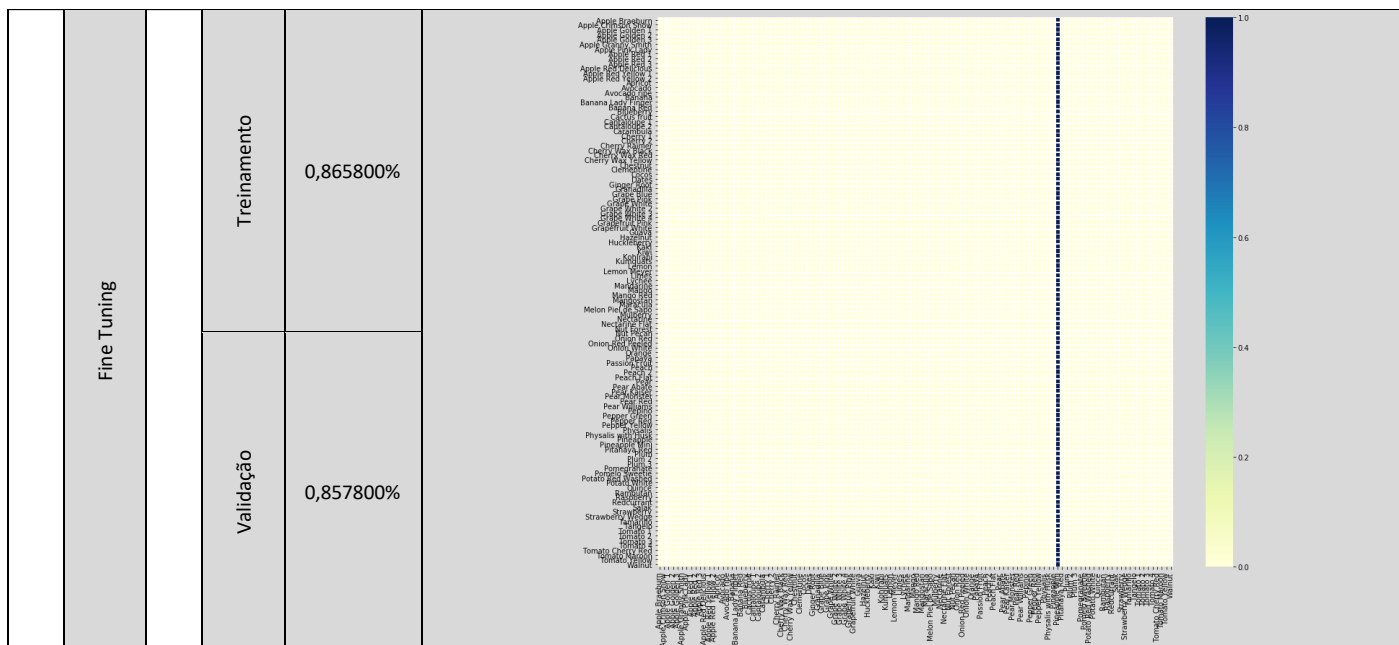
InceptionV3	Feature Extraction	Parâmetros	Treináveis	233.586		
			Não Treináveis	21.802.784		
	Fine Tuning	Parâmetros	Treináveis	22.001.938		
			Não Treináveis	34.432		
MobileNet	Feature Extraction	Parâmetros	Treináveis	1.313.394		
			Não Treináveis	2.257.984		
	Fine Tuning	Parâmetros	Treináveis	3.537.266		
			Não Treináveis	34.112		
ResNet50	Feature Extraction	Parâmetros	Treináveis	2.101.362		
			Não Treináveis	23.587.712		



Acurácia de Treinamento & Validação - Matriz de Confusão da Base Validação



ResNet50				
Feature Extraction		Acurácia		
Fine Tuning		Treinamento	Validação	
		99,873500%	0,840300%	
		98,844500%	0,645600%	
VGG16		Acurácia		
Feature Extraction		Treinamento	Validação	
		99,962900%	97,330200%	



4.3 Resultados Consolidados

Rede	Acurácia		Parâmetros	
	Treino	Validação	Treináveis	Não Treináveis
Homemade	0,999542	0,923319	11.057.34	0
InceptionV3 - Feature Extraction	0,080803	0,083596	233.586	21.802.784
InceptionV3 - Fine Tuning	1	0,994923	22.001.938	34.432
MobileNet - Feature Extraction	0,237961	0,216912	1.313.394	2.257.984
MobileNet - Fine Tuning	1	0,994485	3.537.266	34.112
ResNet50 - Feature Extraction	0,006456	0,008403	2.101.362	23.587.712
ResNet50 - Fine Tuning	0,998735	0,988445	25.635.954	53.120
VGG16 - Feature Extraction	0,999629	0,973302	525.426	14.714.688
VGG16 - Fine Tuning	0,008658	0,008578	15.240.114	0

5. Teste Estatístico

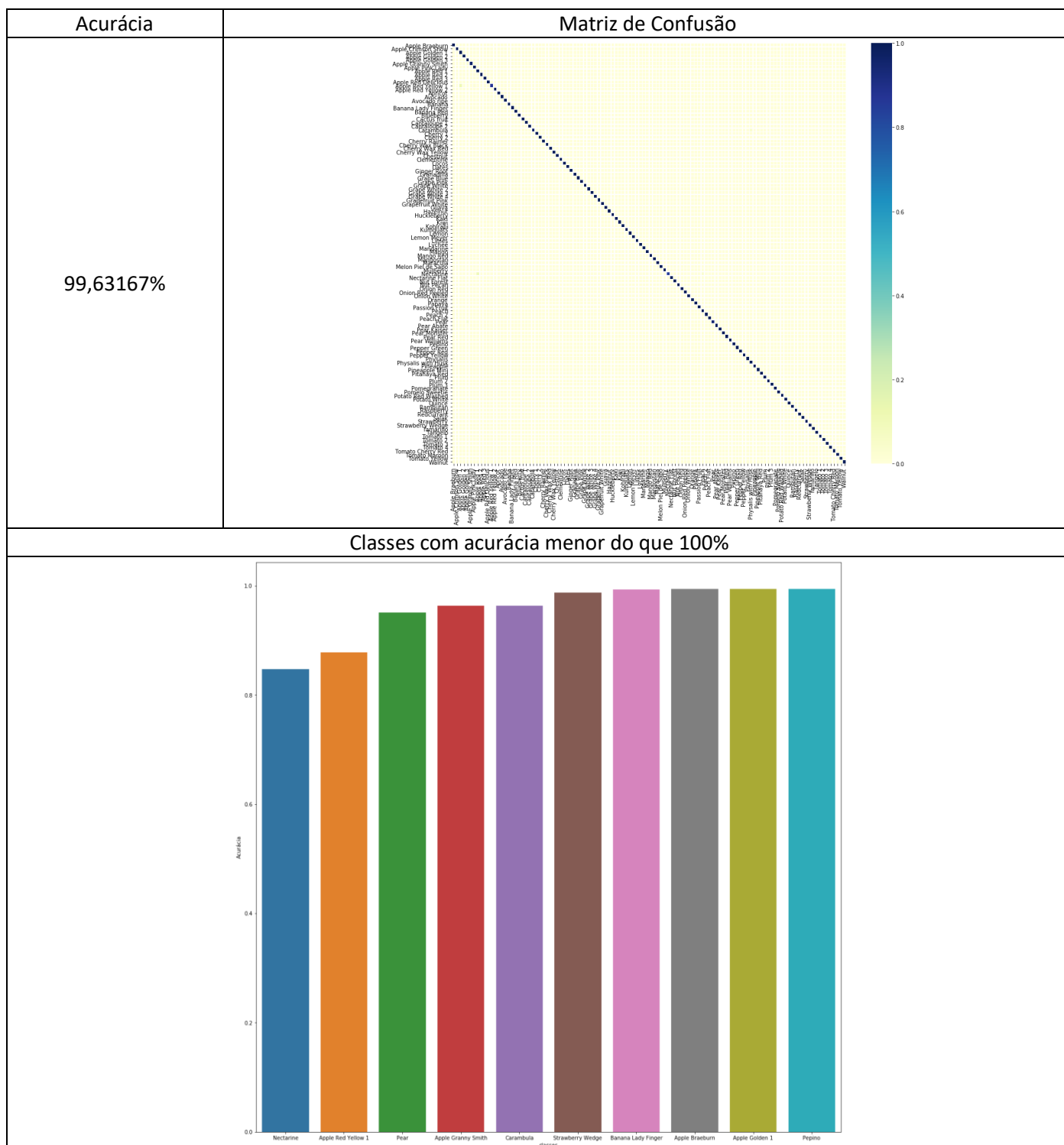
Para a escolha do melhor modelo, foi aplicado um teste estatístico no resultado das redes “InceptionV3 - Fine Tuning”, “MobileNet - Fine Tuning” e “VGG16 - Feature Extraction”. A partir do teste demonstrado abaixo:

	Ganho	P-Valor	Resultado
InceptionV3 x MobileNet	0,07527%	0,79873	Essa diferença não é significativa.
MobileNet x VGG16	2,00475%	1,81616e-05	Essa diferença é significativa.

A rede “MobileNet - Fine Tuning” foi escolhida como melhor modelo e consequentemente foi utilizado para avaliar o conjunto de teste.

6. Aplicando o melhor modelo na base de teste

Ao aplicar o modelo "" na base de Teste, foi obtido os seguintes resultados:



7. Oportunidades de aplicação

7.1 Supermercado

Um caixa onde uma câmera identifica as 5 possibilidades mais próximas e apresenta em tela para que o próprio cliente ou operador de caixa escolha o correto produto trazendo maior agilidade para os clientes no ato de pagar as compras. Adicionalmente, a câmera instalada no caixa tiraria fotos dos produtos gerando assim mais dados para treinamento.

7.2 Acessibilidade

Provido de um aparelho celular com câmera, o deficiente visual aponta a câmera para a gôndola e celular fala as frutas que foram capturadas pela câmera do celular.