UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Allan Cerentini

IDENTIFICAÇÃO DO GLAUCOMA EM IMAGENS DO FUNDO DO OLHO UTILIZANDO APRENDIZAGEM PROFUNDA

Santa Maria, RS
2018

Allan Cerentini

IDENTIFICAÇÃO DO GLAUCOMA EM IMAGENS DO FUNDO DO OLHO UTILIZANDO APRENDIZAGEM PROFUNDA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação, Área de Concentração em Ciência da Computação, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Ciência da Computação**.

ORIENTADOR: Dr. Daniel Welfer

Santa Maria, RS
2018

Cerentini, Allan
Identificação do Glaucoma em Imagens do Fundo do Olho Utilizando Aprendizagem Profunda / Allan Cerentini.- 2018.
84 p.; 30 cm

Orientador: Daniel Welfer
Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-Graduação em Ciência da Computação, RS, 2018

1. Redes Neurais 2. Aprendizado de Máquina 3. Glaucoma I. Welfer, Daniel II. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

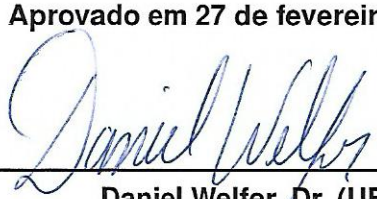
End. Eletr.: acerentini@inf.ufsm.br

Allan Cerentini

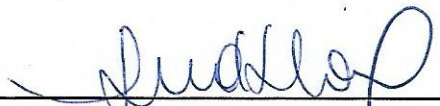
IDENTIFICAÇÃO DO GLAUCOMA EM IMAGENS DO FUNDO DO OLHO UTILIZANDO APRENDIZAGEM PROFUNDA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação, Área de Concentração em Ciência da Computação, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Ciência da Computação.**

Aprovado em 27 de fevereiro de 2018:



Daniel Welfer, Dr. (UFSM)
(Presidente/Orientador)



Andrea Schwertner Charão, Dra. (UFSM)

Fábio Paulo Basso, Dr. (UNIPAMPA)

Santa Maria, RS
2018

AGRADECIMENTOS

Agradeço ao meu pai e a minha mãe por TUDO.

RESUMO

IDENTIFICAÇÃO DO GLAUCOMA EM IMAGENS DO FUNDO DO OLHO UTILIZANDO APRENDIZAGEM PROFUNDA

AUTOR: Allan Cerentini
ORIENTADOR: Daniel Welfer

O Glaucoma é uma doença que danifica o nervo óptico podendo causar perda de visão ou cegueira total. Essa doença é a maior causadora de cegueira irreversível no mundo. Estima-se que até 2020 a quantidade de pessoas com glaucoma poderá chegar em 76 milhões. Este trabalho compara a acurácia de diferentes arquiteturas de redes neurais que utilizam o aprendizado profundo para o reconhecimento de imagens. Essas redes neurais podem auxiliar os profissionais da área de saúde a realizarem o diagnóstico do glaucoma de uma forma mais eficiente e precisa, uma vez que o processo é feito manualmente pelos especialistas. Esse trabalho utiliza um sistema de detecção de objetos do estado da arte com grande performance, chamado de YOLO9000, responsável por detectar o nervo óptico, que é a região de interesse. Após a detecção dessa região é utilizado um classificador, uma rede neural convolucional, para detectar a presença do glaucoma. Esse trabalho analisa diferentes classificadores para verificar qual possui melhor acurácia para a resolução desse problema. Foram utilizados bancos públicos de imagens do fundo do olho para a validação desse processo. A rede neural convolucional chamada de DenseNet foi a que obteve melhor acurácia média na detecção do glaucoma nos bancos de imagens utilizados. Os resultados obtidos foram 100%, 85,8%, 94,4% nos bancos de imagem HRF, RIM-ONE-R1-R2 e RIM-ONE-R3, respectivamente, utilizando a métrica da área da curva de Característica de Operação do Receptor.

Palavras-chave: Redes neurais. Glaucoma. Aprendizado de Máquina.

ABSTRACT

IDENTIFYING GLAUCOMA IN FUNDUS IMAGES USING DEEP LEARNING

AUTHOR: Allan Cerentini
ADVISOR: Daniel Welfer

Glaucoma is a disease that damages the optic nerve and can cause loss of vision or total blindness. This disease is the major cause of irreversible blindness in the world. It is estimated that by 2020 the number of people with glaucoma could reach 76 million. This work compares the accuracy of different neural network architectures that use deep learning for image recognition. These neural networks can help healthcare professionals to diagnose glaucoma more efficiently and precisely, since the process is done manually by specialists. This work utilizes a state-of-the-art, high-performance object detection system called the YOLO9000, responsible for detecting the optic nerve, which is the region of interest. After detection of this region, a convolutional neural network was used to detect the presence of glaucoma. This work analyzes different classifiers to verify which one has the best accuracy to solve this problem. Public available fundus images databases were used to validate this process. The convolutional neural network called DenseNet was the one with best average accuracy to detect the glaucoma among the used images databases. The results were 100 %, 85.8 94.4 % in the HRF, RIM-ONE-R1-R2 and RIM-ONE-R3 image databases, respectively, using the area under the receiver operating characteristic metric.

Keywords: Neural networks. Glaucoma. Machine learning.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

RI	Região de Interesse
RNA	Redes Neurais Artificiais
RNC	Redes Neurais Convolucionais
$ReLU$	Rectified Linear Unit
ULR	Unidade Linear Retificada
$ILSVRC$	ImageNet Large Scale Visual Recognition Challenge
$R-CNN$	Region-based Convolutional Neural Network
$YOLO$	You Only Look Once
VP	Verdadeiro Positivo
VN	Verdadeiro Negativo
FP	Falso Positivo
FN	Falso Negativo
COR	Característica de Operação do Receptor
RGB	Red Green Blue
HRF	High Resolution Backgrounds
BTH	Beijing Tongren Hospital
$MedInfo$	International Medical Informatics Conference
PIL	Python Imaging Library

SUMÁRIO

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

A medicina busca cada vez mais aumentar a precisão do diagnóstico de doenças. Uma doença detectada em seu estágio inicial tem chances muito maiores de ser tratada com sucesso, aumentando a qualidade de vida do paciente e economizando recursos públicos. O uso de imagens digitais em exames trouxe grandes avanços para a medicina. Essa adoção permitiu que milhares de exames pudessem ser armazenados e compartilhados de forma eficiente, também possibilitando o desenvolvimento de ferramentas digitais para o auxílio de diagnóstico.

Uma técnica que utiliza imagens digitais é a aprendizagem de máquina, a qual permite com que o computador aprenda a classificar imagens. Para que isso seja possível o computador precisa analisar várias imagens de determinado exame, devidamente classificadas por especialistas, afim de poder classificar novas imagens. O uso da inteligência artificial, área na qual o aprendizado de máquina faz parte, na saúde irá crescer em dez vezes nos próximos cinco anos (Eduardo Prado, 2016). Isso possibilitará com que exames sejam realizados de forma mais precisa e rápida. O diagnóstico das doenças cujo os exames geram imagens digitais, como o glaucoma e câncer, serão os mais beneficiados.

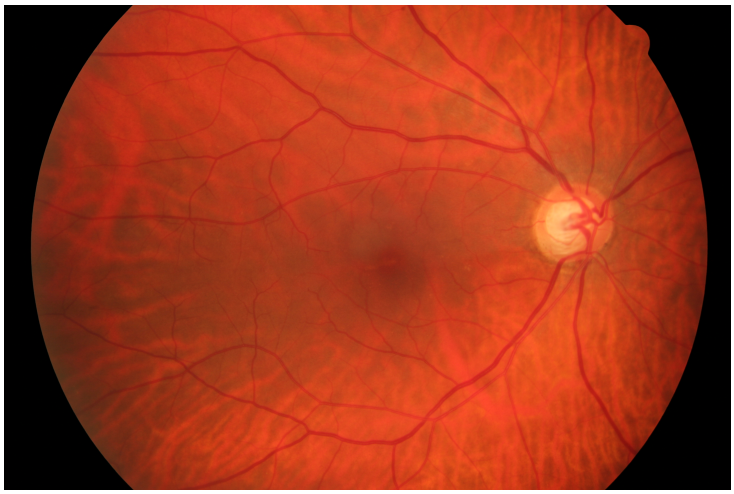
No que corresponde ao diagnóstico do glaucoma, já existem métodos que utilizam imagens digitais para esse fim, auxiliando o diagnóstico precoce dessa doença. De acordo com National Eye Institute (2014), glaucoma é um grupo de doenças que ocorrem nos olhos sendo capaz de danificar o nervo óptico, causando perda de visão parcial ou total. Essa doença, junto com a catarata e o tracoma, são as maiores causadoras de cegueira no mundo, sendo o glaucoma a líder em causar cegueira irreversível (QUIGLEY, 2006). A pesquisa realizada por (THAM et al., 2014) afirma que em 2013, 64,3 milhões de pessoas com idades entre 40 a 80 anos possuíam glaucoma. O mesmo estudo aponta projeções para 2020 e 2040, onde esse número vai aumentar em 18,3% até 2020, elevando a essa quantidade para 76 milhões de pessoas. Em 2040 acontecerá um aumento de 74% comparado a 2013, elevando esse número para 111,88 milhões de pessoas. Em 2020 o total estimado de indivíduos com glaucoma somente na América Latina vai ser de aproximadamente 8 milhões.

Essa doença não pode ser facilmente detectada em suas primeiras etapas. O trabalho realizado por (WILSON, 2009) apresenta as fases dessa doença. Em seu primeiro estágio ela é indetectável e mata lentamente as células do gânglio retinal, responsáveis por receber informações visuais dos fotorreceptores. Em seu próximo estágio ela é assintomática, mas pode ser detectada. O sintoma mais comum neste estágio é a diminuição da

visão periférica, muitas vezes não percebida pelo paciente. Em seu último estágio, danos severos no nervo óptico causam perdas irreversíveis de visão.

De acordo com (Glaucoma Research Foundation, 2017), um exame completo do olho é composto por cinco testes: tonometria, fundoscopia, perimetria, gonioscopia e paquimetria. Focaremos no teste de fundoscopia, pois este gera uma imagem digital, como a da Figura 1.1.

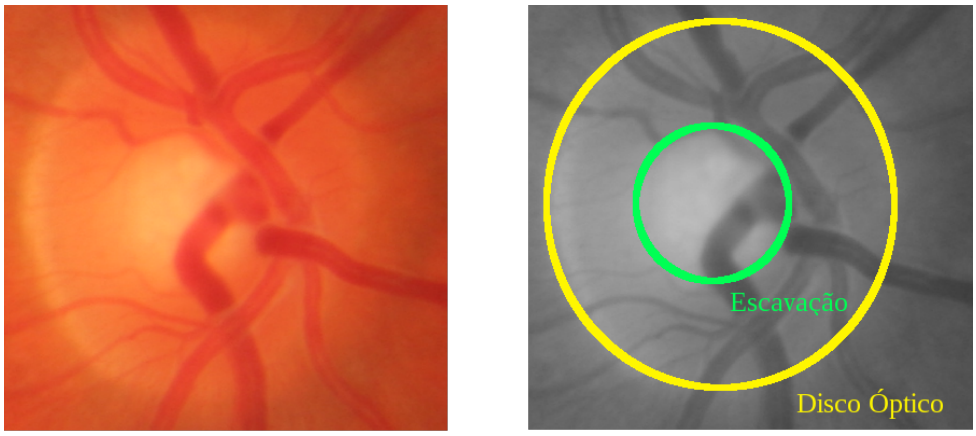
Figura 1.1 – Imagem do fundo do olho, gerada por um retinógrafo.



Fonte: Retirado do banco de imagens de (KOLAR et al., 2013).

Desta imagem é extraída a região de interesse (RI) utilizada neste trabalho, como ilustrada na Figura 1.2. Essa região contém o disco óptico e a escavação. A escavação é uma área natural do nervo óptico, localizada no centro do mesmo, que não possui fibras nervosas, ocasionando um ponto cego. O tamanho deste ponto cego varia de pessoa para pessoa, mas o quanto esse ponto cego está aumentando em relação ao tamanho original pode ser um indicativo da presença do glaucoma. Essa mudança de tamanho está muitas vezes relacionada ao aumento da pressão interna do olho.

Figura 1.2 – RI, composta pelo disco óptico e escavação. Em a) temos a RI extraída de uma imagem do fundo do olho. Em b) temos a imagem a) em escala de cinza para melhor visualização, assim como marcações aproximadas da região do disco óptico e do nervo óptico.



a) b)

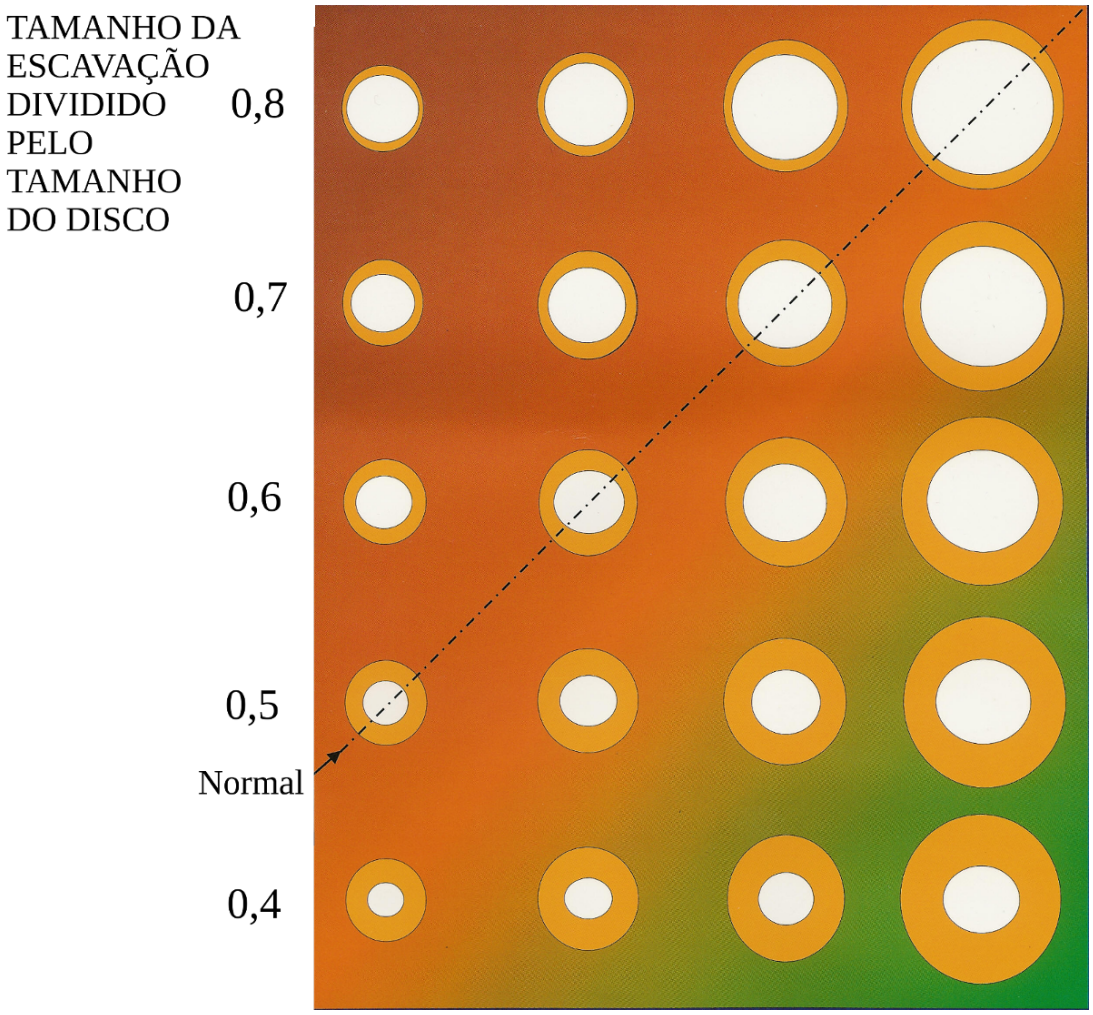
Fonte: Adaptada do banco de imagens (KOLAR et al., 2013).

Essa RI é normalmente analisada tendo como referência a razão entre o tamanho da escavação e o tamanho do disco óptico. A Figura 1.3 apresenta um guia para o especialista detectar a presença do glaucoma em imagens do fundo do olho, de acordo com razão entre o tamanho da escavação com o tamanho do disco óptico. Esse processo é feito manualmente e demanda tempo. O trabalho proposto faz um comparativo de diversos modelos de RNC, afim de verificar qual apresenta a melhor acurácia para futuros sistemas que auxiliam a detecção do glaucoma. Para que um sistema possa detectar a presença do glaucoma é preciso:

- Detectar a RI: O sistema precisa detectar e extrair automaticamente a RI. Para esse fim, ele precisa ser treinado para diferenciar a RI de outras estruturas do fundo do olho.
- Classificar a RI: Com o intuito que o sistema consiga classificar essa região, ele deverá ser treinado em imagens anotadas por especialistas, com o propósito de classificar imagens ainda não processadas pelo mesmo.

Após treinado, utilizando bancos de imagens públicas, o sistema deverá ser capaz de detectar a presença do glaucoma em imagens ainda não vistas pelo mesmo.

Figura 1.3 – Tabela indicando o tamanho normal de uma escavação, em branco, em relação ao tamanho do disco óptico, em laranja. Abaixo da linha tracejada se encontram as proporções de escavação consideradas aceitáveis pelo tamanho do disco óptico. Acima dessa linha há suspeita de glaucoma.

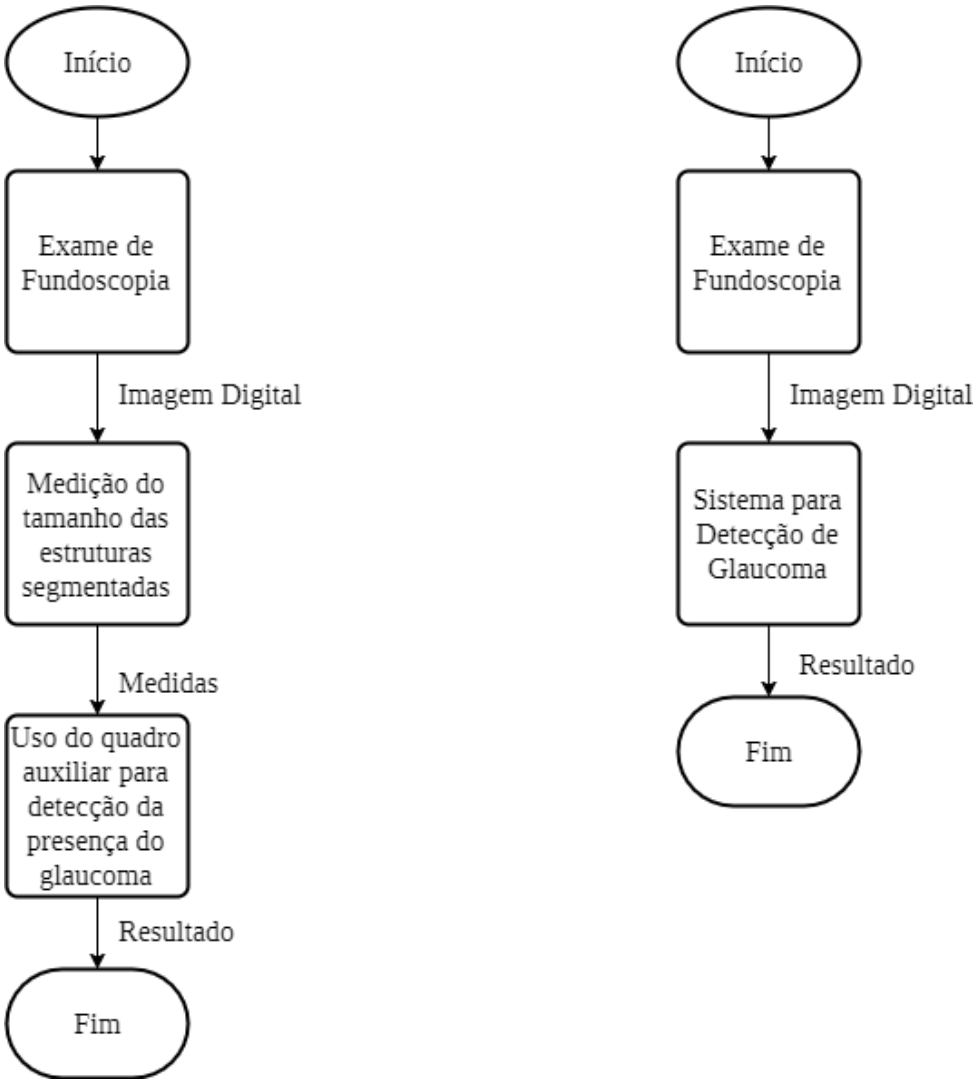


Fonte: Adaptado de Cheshire Local Optical Committee [1].

A Figura 1.4 compara os processos da detecção do glaucoma. Em a) o processo é realizado manualmente por um especialista e em b) o processo realizado pelo sistema proposto neste trabalho. Para a detecção manual, processos adicionais são necessários. As medidas das estruturas são realizadas de forma cuidadosa, pois cada milímetro é decisivo para a identificação de um glaucoma em seu estágio inicial, tornando o processo vagaroso. Mas no processo manual também é verificada a intensidade do glaucoma. O comparativo proposto tem como objetivo somente detectar a presença do mesmo, caso positivo, exames adicionais devem ser realizados para maiores detalhes.

[1] Disponível em: <http://www.loc-net.org.uk/cheshire/useful-documents/documents/disc-size-chart/> Acesso em jan. 2018.

Figura 1.4 – Diagrama comparando os processos de detecção do glaucoma. Em a) o processo é realizado manualmente pelo especialista. Em b) o processo é realizado automaticamente utilizando um sistema com redes neurais, diminuindo a quantidade de etapas manuais, acelerando o processo de detecção do glaucoma.



Fonte: Autoria própria.

1.2 OBJETIVOS

Este trabalho tem o objetivo geral realizar um comparativo de diversos modelos de RNC, para que futuros sistemas que auxiliam o exame do glaucoma possam utilizar.

Como objetivos específicos, este trabalho visa:

- Utilizar bancos de imagens públicas para treinar e validar o sistema.
- Adaptar um sistema de detecção de objetos genéricos para detectar a RI.

- Treinar e validar diversas arquiteturas de redes neurais, classificadoras de imagens, para detectar a presença do glaucoma
- Avaliar qual arquitetura apresenta a maior acurácia para esse problema.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está organizada como segue. O Capítulo 2, exibe a fundamentação teórica, explicando o que são redes neurais e um breve resumo sobre cada arquitetura utilizada neste trabalho, assim como sistemas de detecção de objetos. O Capítulo 3 descreve os trabalhos relacionados. O Capítulo 4, descreve o desenvolvimento do método utilizado para detecção e classificação, assim como os experimentos realizados para decidir a melhor arquitetura e ferramentas utilizadas para o auxilio. O Capítulo 5, demonstra o resultado de todos os experimentos realizados assim como gráficos do processo de treinamento. O Capítulo 6, apresenta a conclusão do trabalho e trabalhos futuros.

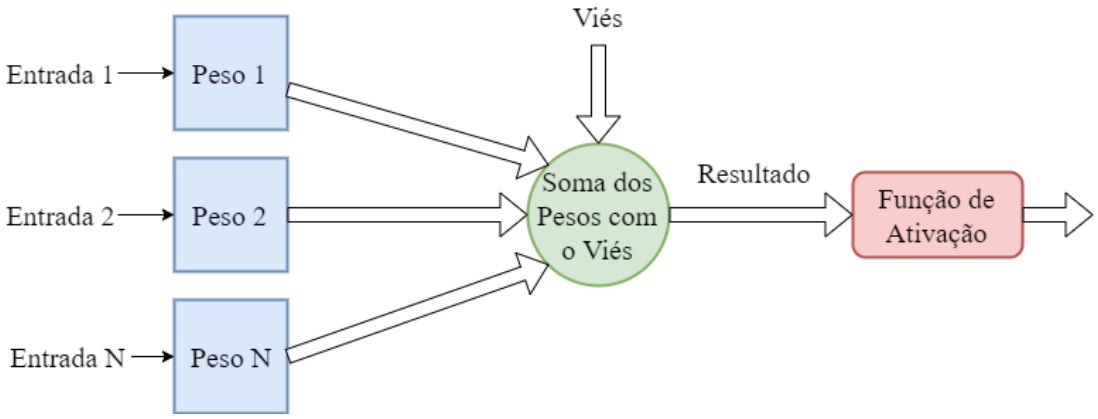
2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta informações sobre redes neurais, que foram utilizadas para o desenvolvimento de um classificador de imagens, e o detector de objetos. A seção 2.1 fornece uma introdução sobre o que é uma rede neural artificial. A seção 2.2 explica as características básicas da rede neural convolucional. A seção 2.3 apresenta algumas das arquiteturas desenvolvidas para a classificação de imagens. A seção 2.4 apresenta o funcionamento de redes neurais utilizadas para localizar objetos em imagens.

2.1 REDES NEURAIS E APRENDIZADO PROFUNDO

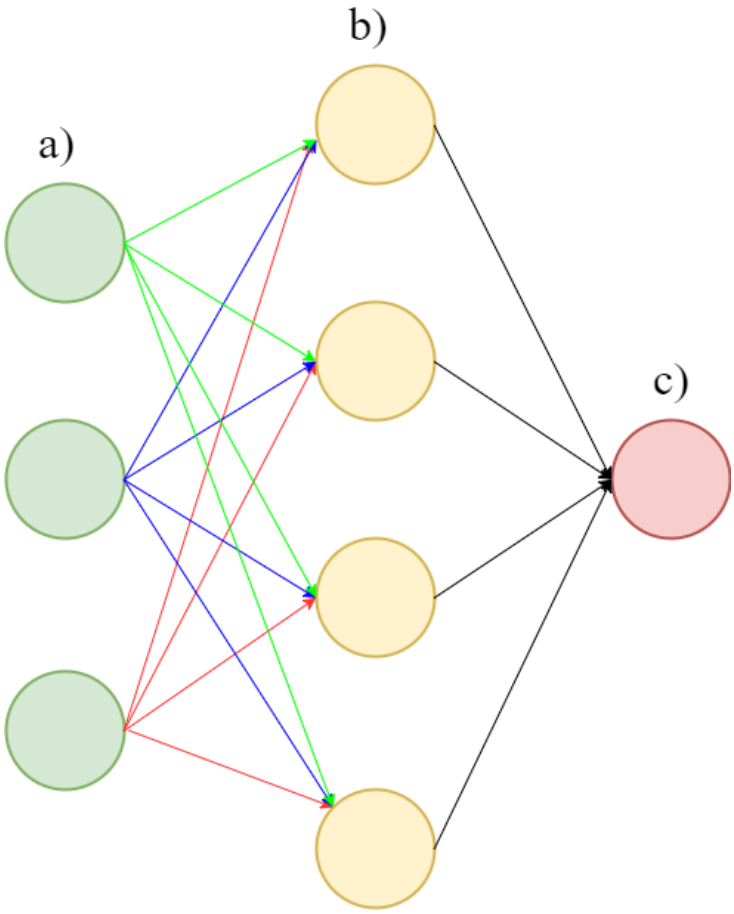
O aprendizado de máquina é um subcampo da área da ciência da computação, responsável por desenvolver sistemas capazes de auxiliar tomadas de decisões, resolver problemas e realizar previsões a partir do uso dos dados disponíveis. As Redes Neurais Artificiais (RNA) fazem parte do conjunto de técnicas utilizadas nessa área. Estas por sua vez foram inspiradas no processo de aprendizagem do cérebro biológico. Formadas por um conjunto de neurônios, como os da Figura 2.1 interligados. Essas ligações são feitas de forma que a saída de um neurônio seja a entrada de outro, mas nunca de maneira que possa gerar um ciclo infinito. Esses neurônios são organizados em forma de camadas, de maneira que a entrada seja processada por conjuntos paralelos de neurônios. Cada neurônio de uma camada é conectado com todos os neurônios da camada anterior, como pode ser visto na Figura 2.2. Quando uma RNA apresenta várias camadas é utilizado o termo "aprendizagem profunda"para se referir a arquitetura. A Figura 2.1 apresenta as principais estruturas de um neurônio artificial. Cada neurônio recebe diversos sinais de entrada, que podem vir dos dados de entrada ou de outros neurônios. Cada sinal recebido é multiplicado por um valor chamado de peso. Todos os sinais de entrada são multiplicados pelos seus respectivos pesos e seus resultados são somados. Um viés, que é um valor extra, também é somado com o valor total, com o propósito de aumentar o grau de liberdade de ajuste dos pesos na fase de treinamento. A soma total é enviada para uma função de ativação, a qual verifica se o resultado final é suficiente para ser enviado adiante.

Figura 2.1 – Principais estruturas de um neurônio artificial.



Fonte: Adaptada de (Anderson Vinicius, 2017).

Figura 2.2 – Neurônios de uma camada se ligam a todos os outros neurônios das próxima camada. a) a camada de entrada. b) a camada intermediária. c) a camada de saída.



Fonte: Autoria própria.

O processo de aprendizagem de uma RNA é realizado através do ajuste dos pesos e do viés de cada neurônio. Essas redes geralmente são treinadas utilizando o processo

de aprendizagem supervisionada, no qual os dados fornecidos para a rede já possuem um rótulo ou resultado esperado. Segundo (Ujjwal Karn, 2016) geralmente o processo de treinamento de uma rede neural ocorre através das seguintes etapas:

- Os valores inicias de todos os pesos da rede geralmente são iniciados de forma aleatória. Isso ocorre somente uma vez e quando a rede é utilizada para o treino.
- A rede recebe uma entrada de dados, com as dimensões previamente adaptadas para a entrada da rede. Então é realizado o processo de propagação, que consiste em processar a informação de entrada utilizando os pesos da rede. O resultado final é utilizado para verificar a probabilidade da entrada pertencer a um dos rótulos disponíveis.
- A diferença entre o valor obtido e o valor esperado é calculado, através de uma função de custo.
- É utilizada a técnica de retro propagação para calcular a contribuição de cada neurônio com o erro. Então a técnica do gradiente descendente é utilizada para atualizar os valores dos pesos para que o erro seja minimizado (SCHMIDHUBER, 2015). Pesos que mais influenciaram a proporção desse erro receberão ajustes maiores.
- As etapas anteriores são repetidas para cada entrada. Cada entrada contribui para o ajuste de pesos, até que um valor médio ideal seja encontrado, e a maioria das entradas possam ser classificadas corretamente.

A arquitetura de uma rede neural é geralmente formada por camadas de três tipos: camada de entrada, camadas intermediárias e camadas de saída. A camada de entrada é responsável por receber os dados que serão processados pela rede, como: números, letras, sinais de áudio, pixeis de imagens. As camadas intermediárias são responsáveis por guardar os pesos calculados pela rede em seu processo de treino. A camada de saída é responsável por calcular a probabilidade da entrada pertencer a uma das classes, utilizando o valor processado pelas camadas intermediárias.

De acordo com (BENGIO; COURVILLE; VINCENT, 2013) uma RNA com múltiplas camadas intermediárias conseguem guardar mais informações, possibilitando trabalhar com entradas mais complexas, como imagens. Outro fator importante foi a quantidade de dados públicos digitais disponíveis, pois para se obter uma boa acurácia é necessário grande quantidade de dados.

2.2 REDES NEURAIS CONVOLUCIONAIS

O computador interpreta imagens como um conjunto de números; uma imagem colorida, com 3 canais de cores, com dimensões de 250×250 é vista como uma matriz de 187.500 números para o computador. RNAs ao processarem imagens utilizam essa matriz de números como entrada, na qual cada número pode variar de 0 até 255, representando a intensidade de cada pixel. Uma vez que cada neurônio de uma camada é conectado a todos os neurônios da camada anterior, isso representa uma quantidade de 187.500 pesos para a rede ajustar, somente em uma camada.

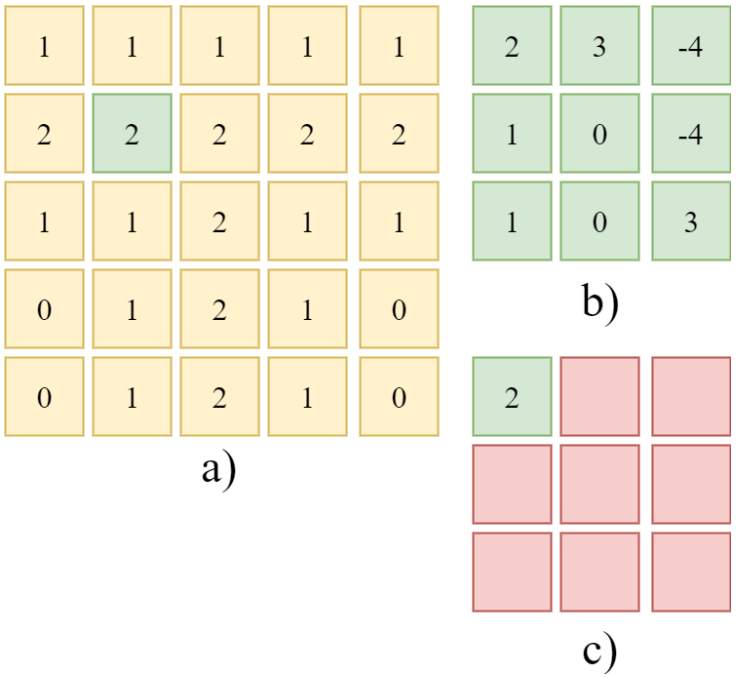
Redes Neurais Convolucionais (RNC) são RNA adaptadas para lidarem com imagens de uma forma mais eficiente. Essa adaptação também teve como inspiração o modo com que a natureza lida com esse problema, principalmente os seres humanos que possuem facilidade em reconhecer pessoas e objetos (MATSUGU et al., 2003). O córtex visual, responsável por essa habilidade, é formado por um complexo conjunto de neurônios, os quais reagem apenas em pequenas regiões do campo visual. Essas sub-regiões são chamadas de campo receptivo. O conjunto de todos os campos receptivos desses neurônios formam o campo visual. Ao utilizar uma técnica similar é possível reduzir drasticamente a quantidade de parâmetros necessários para a atualização da rede.

Segundo (Ujjwal Karn, 2016) uma RNC é formada basicamente por quatro tipos de operações: convolução, ReLU, *pooling* e classificação.

2.2.1 Convolução

As camadas de convolução tem como objetivo aprender características relevantes sobre a imagem, de forma eficiente. Para que isso seja possível é utilizada a técnica de conectividade local, a mesma utilizada pelo córtex cerebral, na qual o campo receptivo do neurônio tem o nome de filtro. Esse filtro é responsável por processar a imagem ou a camada anterior, realizando a soma do produto de todos os valores de seu campo receptivo. Esse processo ocorre em todas as posições de sua entrada, resultando em uma operação com o nome de convolução. A Figura 2.3 apresenta esse processo: o filtro b) de tamanho 3×3 percorre todos os possíveis pontos de a) e marca seu resultado em c). Os valores de b) são ajustados pela rede a fim de possibilitar mapas de característica mais relevantes para as próximas camadas.

Figura 2.3 – Processo de convolução de uma RNC. Em a) temos uma entrada, que pode ser uma imagem ou um mapa de características. Em b) temos o filtro sendo aplicado ao pixel marcado em verde em a). Em c) o mapa de características e marcado em verde o resultado do filtro.



Fonte: Autoria própria.

O filtro recebe uma matriz com 3 dimensões para realizar o processo, formada por: altura, largura e profundidade, no qual a largura e a altura representam o tamanho do campo receptivo e a profundidade o número de canais da imagem ou camada anterior. Geralmente somente o tamanho do campo receptivo é ajustado. O filtro sempre utilizará todos os canais da mesma localidade para realizar as operações (Andrej Karpathy, 2017). Uma vez que o filtro é o mesmo para todos os neurônios formando o campo visual, é possível detectar características independente de sua posição na imagem.

A convolução dos filtros nas entradas da rede geram mapas de características, que armazenam os locais onde o filtro detectou uma característica importante na imagem. A quantidade desses mapas é dada pela quantidade de diferentes filtros que formam a camada convolucional. Cada neurônio utilizando o mesmo filtro marca a intensidade no mesmo mapa, de acordo com a probabilidade de tal característica estar presente em sua região. Filtros mais próximos da entrada da rede tendem a guardar características mais simples como: cantos, linhas, curvas. Esses mapas são utilizados pelas próximas camadas que realizam o mesmo processo, mas desta vez os filtros buscarão por características mais complexas como: formas geométricas. Esse processo ocorre até ao final da rede, que contém representações próximas do objeto buscado pela rede.

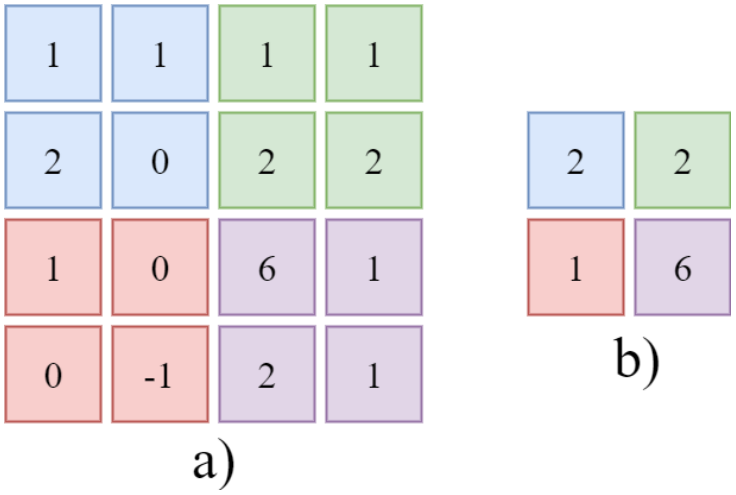
2.2.2 ReLU

O resultado final da convolução é processado por uma função ativação, onde que a mais utilizada é a de Unidade Linear Retificada (ULR), também conhecida por ReLU (*Rectified Linear Units*). Essa função é responsável por substituir todos os valores menores que zero em zero, introduzindo não linearidade a rede, uma vez que a rede tem o propósito de resolver problemas não lineares (Ujjwal Karn, 2016).

2.2.3 Pooling

A camada de *pooling* é responsável por reduzir dimensionalmente cada mapa de característica, mantendo somente as informações mais relevantes. Essa técnica consiste em deslizar um filtro, sem ele sobrepor regiões, com tamanho X por toda a área do mapa e aplicar uma função que condensará a informação da região observada pelo filtro. A técnica mais comum é a de *max pooling*, que utiliza o valor máximo encontrado na região como resultado. A Figura 2.4 demonstra o funcionamento do algoritmo de *max pooling*, as cores de a) representam agrupamentos montados pelo algoritmo, o valor máximo de cada agrupamento é utilizado é armazenado em b).

Figura 2.4 – Processo de *max pooling* de uma RNC. Em a) temos uma entrada, que pode ser uma imagem ou um mapa de características. Em b) temos o resultado do processo de *max pooling*.



Fonte: Autoria própria.

Essa camada é responsável aumentar a robustez da rede em relação a distorções. Pequenas transformações e distorções prejudicam menos a acurácia da rede, uma vez que geralmente a camada de *pooling* escolhe os valores mais altos da região. Isso favorece a detecção em imagens de diferentes escalas devido a concentração de informação.

Reduz o a quantidade de parâmetros da rede, aumentando sua performance e reduzindo o sobreajuste.

O sobreajuste ocorre quando temos muitos parâmetros para pouca quantidade de dados, então a rede começa a decorar as entradas em vez de aprender características gerais. A rede com sobreajuste apresenta grande acurácia durante o treinamento, mas baixa acurácia quando testada em entradas diferentes das utilizadas em sua fase de treino. A redução de parâmetros realizada pela camada de *pooling* melhora a capacidade de generalização da rede.

2.2.4 Classificação

A última camada da rede é responsável por classificar a entrada em uma das classes disponíveis. Essa camada geralmente é composta por neurônios inteiramente conectados, os mesmos vistos em redes neurais comuns, que interpretam os últimos mapas de característica da rede. Essa interpretação ocorre devido aos neurônios, desta camada, estarem inteiramente conectados com todos os mapas de características da camada anterior, sendo capaz de misturar essas informações a fim de que uma visão geral da imagem possa ser adquirida. Então a função *softmax*, Equação 2.1, geralmente é utilizada para representar a probabilidade de cada classe. A equação utiliza como entrada um vetor x_i que representa a saída de cada neurônio da última camada e como saída um valor para cada entrada, representando a probabilidade, cuja a soma é igual a 1. (RAWAT; WANG, 2017).

$$(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \tag{2.1}$$

2.3 CLASSIFICAÇÃO DE OBJETOS

A classificação de objetos em imagens se tornou uma das áreas em que as redes neurais tiveram destaque. A quantidade de imagens digitais disponíveis aumentou com o passar do tempo, mostrando o ponto fraco de muitas arquiteturas de redes neurais convolucionais. Muitos modelos exigem muito poder computacional, isso fica cada vez mais evidente conforme o tamanho do banco de imagens, tornando a fase de treino custosa.

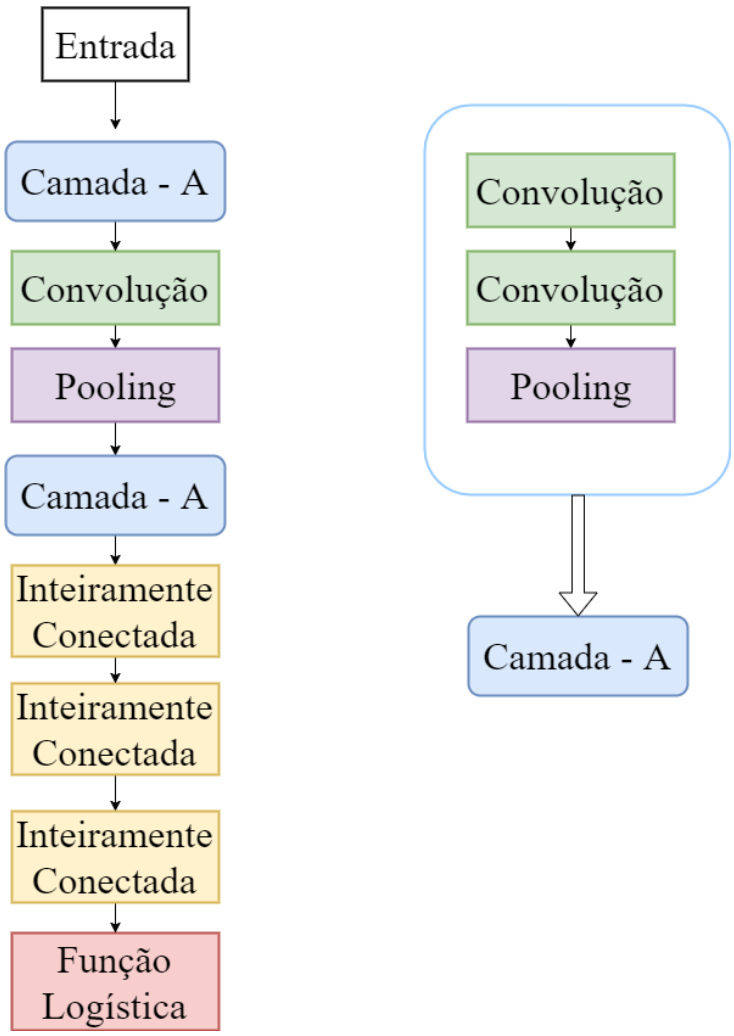
Outros modelos não possuem os melhores parâmetros para lidar com determinados problemas. Então surgiram novas técnicas para se construir arquiteturas de RNCs.

2.3.1 AlexNet

A rede desenvolvida por (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) ganhou o ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2012, sendo a primeira vez que uma rede neural conseguiu o primeiro lugar nessa competição. Essa vitória foi um grande marco, pois mostrou que já existia poder computacional disponível para o uso de redes neurais. Essa vitória serviu de inspiração para o desenvolvimento de novos modelos de redes neurais que foram e ainda são utilizados em competições de classificação de imagens.

A arquitetura da Alexnet, que pode ser vista na Figura 2.5, é formada por 5 camadas convolucionais, com filtros com tamanho até 11×11, e 3 camadas totalmente conectadas, totalizando 8 camadas que guardam pesos. Também eram aplicadas camadas de *max pooling*. As camadas inteiramente conectadas são formadas por neurônios que se conectam a todos os neurônios da camada anterior, como visto na Figura 2.2. Essa sequência de camadas permite com que a rede possa combinar os últimos mapas de características, os que possuem características de alto nível, para representar as classes buscadas.

Figura 2.5 – Rede AlexNet.

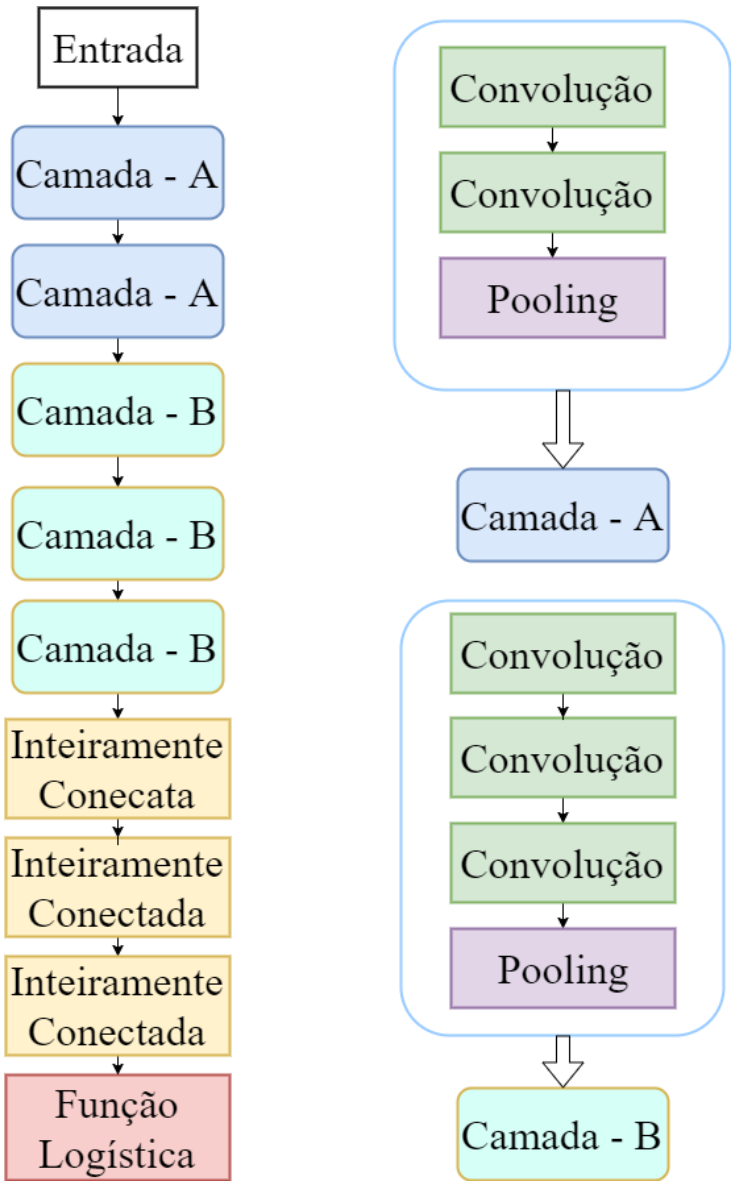


Fonte: Adaptado de (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

2.3.2 VGG

O modelo proposto por (Simonyan; Zisserman, 2014) apresenta camadas simples mas com propósito de ser mais profunda do que a AlexNet. A rede é formada basicamente por filtros com tamanho 3×3 e camadas de *max pooling* para a redução de parâmetros. Foi uma rede que gerou contribuições importantes como: redes mais profundas tendem a possuir maior acurácia, utilizar filtros com tamanho 3×3 em sequência pode substituir filtros maiores e redes profundas inteiramente conectadas são difíceis de serem treinadas. Após testes com diferentes profundidades a versão com 16 camadas, que pode ser vista na Figura 2.6,que guardam os pesos da rede, foi considerada a com melhor desempenho, assim dando o nome de VGG-16.

Figura 2.6 – VGG16.



Fonte: Adaptado de (Simonyan; Zisserman, 2014).

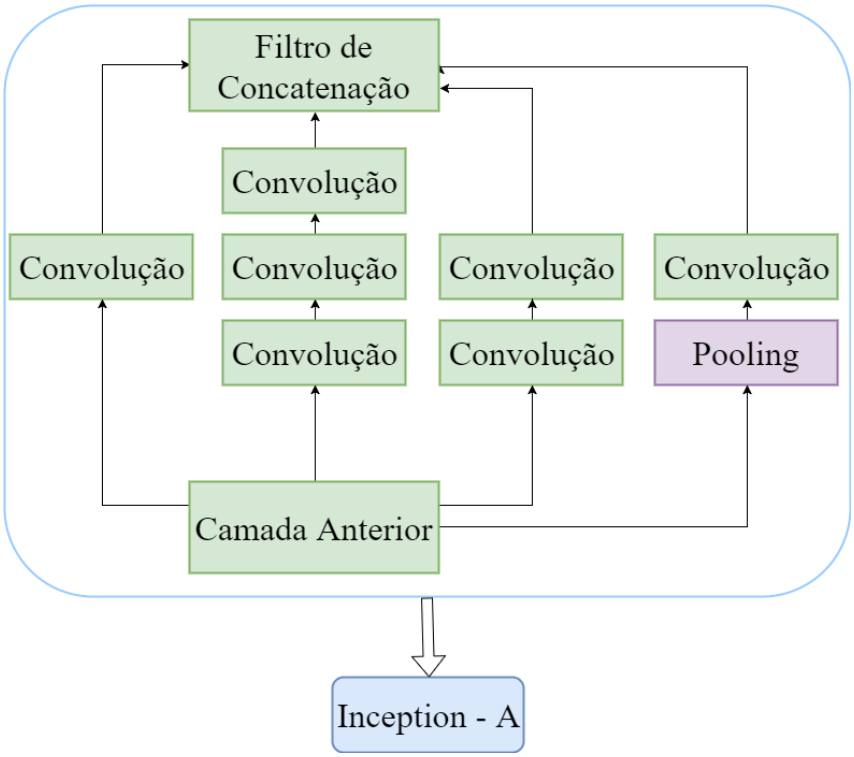
2.3.3 Inception v3

A rede neural desenvolvida por (Szegedy et al., 2015) tem como objetivo diminuir a complexidade, número de parâmetros, e melhorar a eficiência computacional em relação as arquiteturas anteriores. Essa rede desenvolvida pela *Google* ganhou o ILSVRC 2014.

Ao se desenvolver uma rede neural é preciso tomar decisões de qual tamanho de filtro mais eficiente utilizar em cada camada, dentre os mais comuns como: 1×1, 3×3 ou 5×5 . Filtros muito grandes podem acabar por aumentar a complexidade de forma

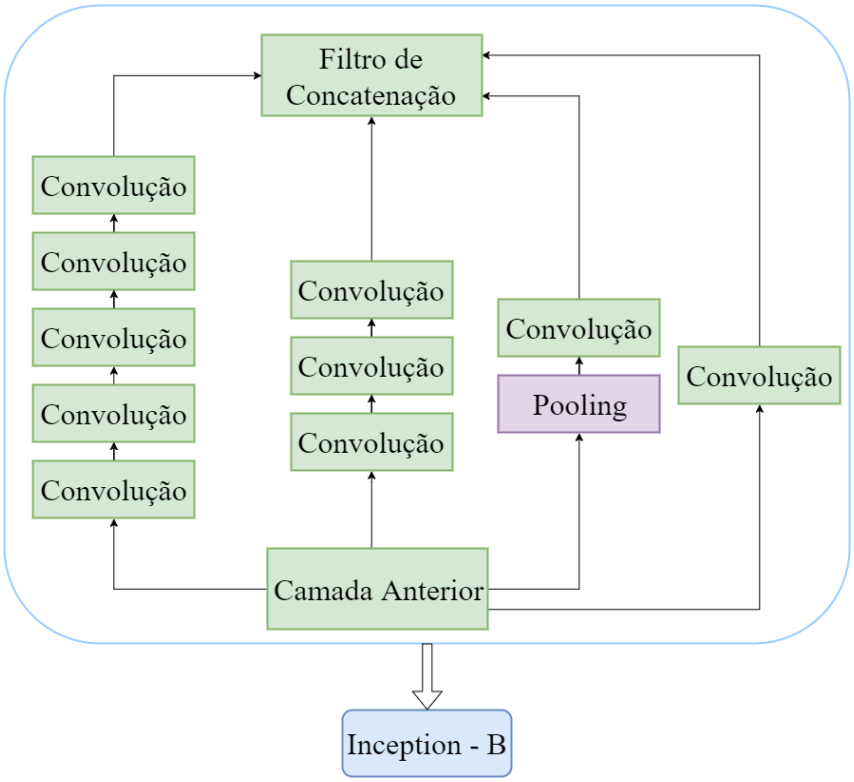
desnecessária, causando *overfitting*, e filtros muito pequenos podem não ser suficientes para capturar a informação necessária. Para resolver esse problema, foi desenvolvido um bloco chamado de *inception*, que pode ser visto na Figura 2.7, Figura 2.8 e Figura 2.9. Esse bloco aplica filtros de tamanho: 1×1, 3×3 e 5×5 em paralelo, onde que esses podem ser precedidos por filtros com tamanho 1×1 ou por *max pooling* para reduzir a complexidade dos parâmetros. Então o resultado final desses filtros é concatenado, permitindo com que a rede possa escolher qual o melhor tamanho de filtro para a resolução do problema. Na sua terceira versão os filtros de tamanho 5×5 foram substituídos por dois filtros de tamanho 3×3 aplicados em sequência. A Figura 2.10 apresenta como essas camadas formam a rede.

Figura 2.7 – Primeira variação da camada *Inception*.



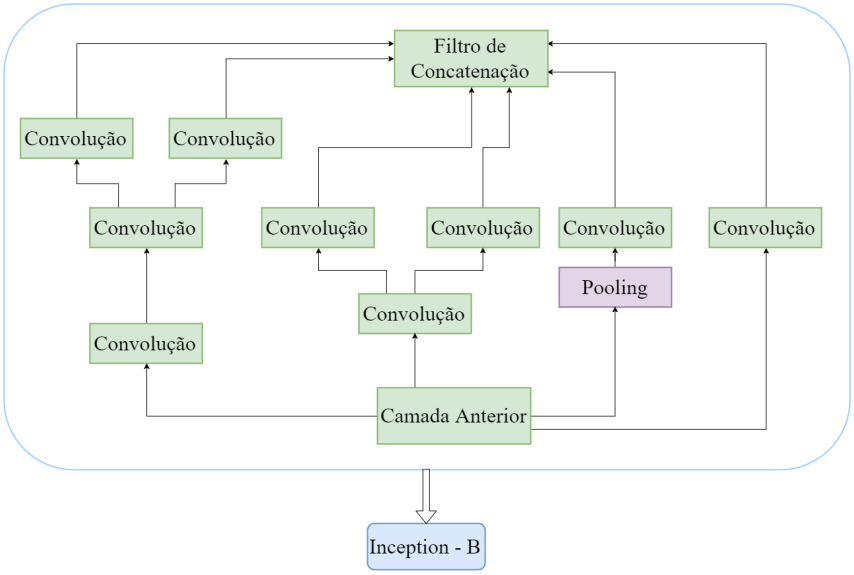
Fonte: Adaptado de (Szegedy et al., 2015).

Figura 2.8 – Segunda variação da camada *Inception*.



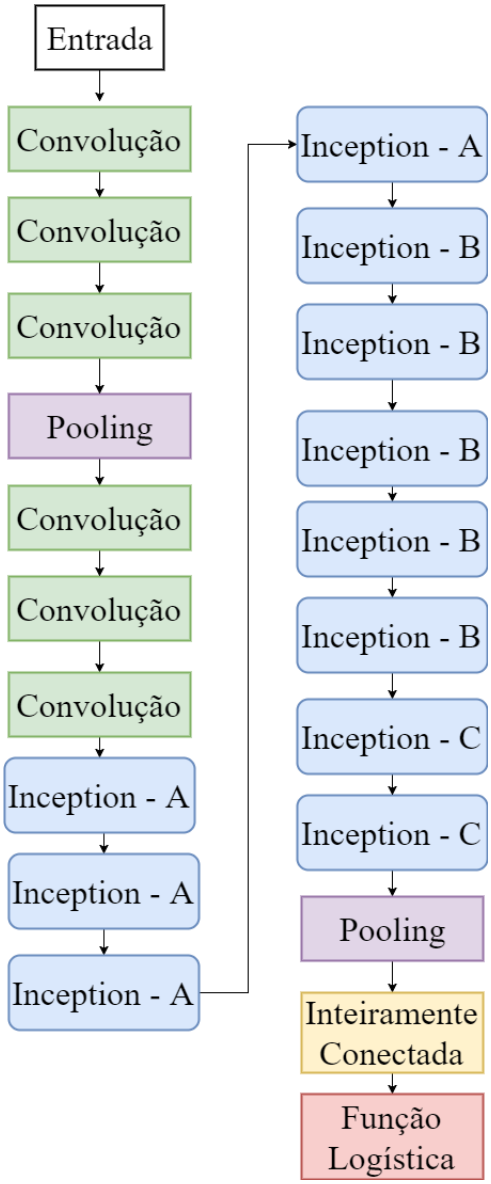
Fonte: Adaptado de (Szegedy et al., 2015).

Figura 2.9 – Terceira variação da camada *Inception*.



Fonte: Adaptado de (Szegedy et al., 2015).

Figura 2.10 – Arquitetura da Inception v3.



Fonte: Adaptado de (Szegedy et al., 2015).

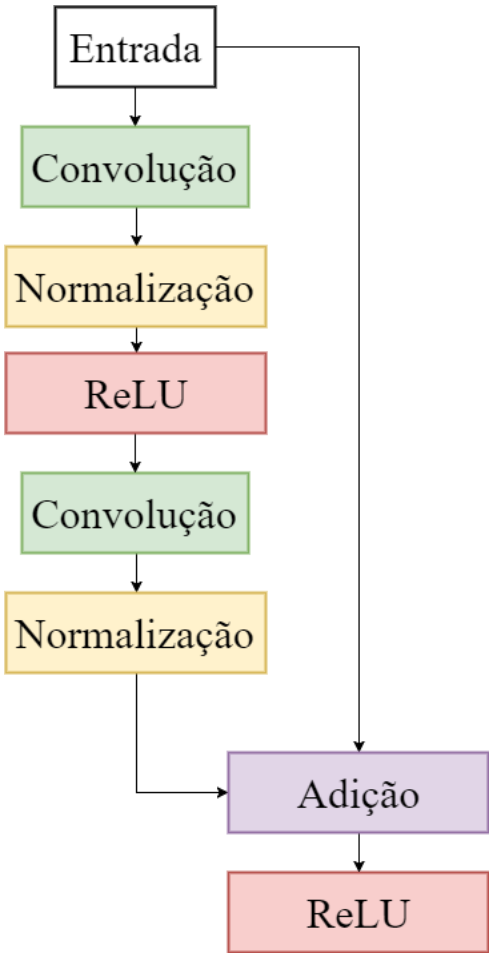
2.3.4 ResNet

O trabalho desenvolvido por (HE et al., 2015) apresenta uma arquitetura com muitas camadas de profundidade em relação as demais e mantendo bom desempenho em seu treino. Essa rede desenvolvida pela equipe da *Microsoft*, composta por 152 camadas, venceu o ILSVRC 2015, com acurácia maior do que seres humanos.

Redes neurais mais profundas geralmente apresentam maior acurácia, mas a performance ao treinar um conjunto de dados diminui, pois camadas muito profundas pos-

suem dificuldades em retro propagar erros, uma vez que a entrada acaba sendo degradada nas últimas camadas. Para resolver esse problema foi desenvolvido o bloco residual, como visto na Figura 2.11. Esse bloco processa uma entrada através de várias camadas de pesos e então o resultado desse processo é adicionado à entrada. Isso permite que algumas camadas que não consigam aprender características relevantes possam ser ignoradas, preservando características importantes para as camadas mais profundas.

Figura 2.11 – Bloco residual que compõe a ResNet.



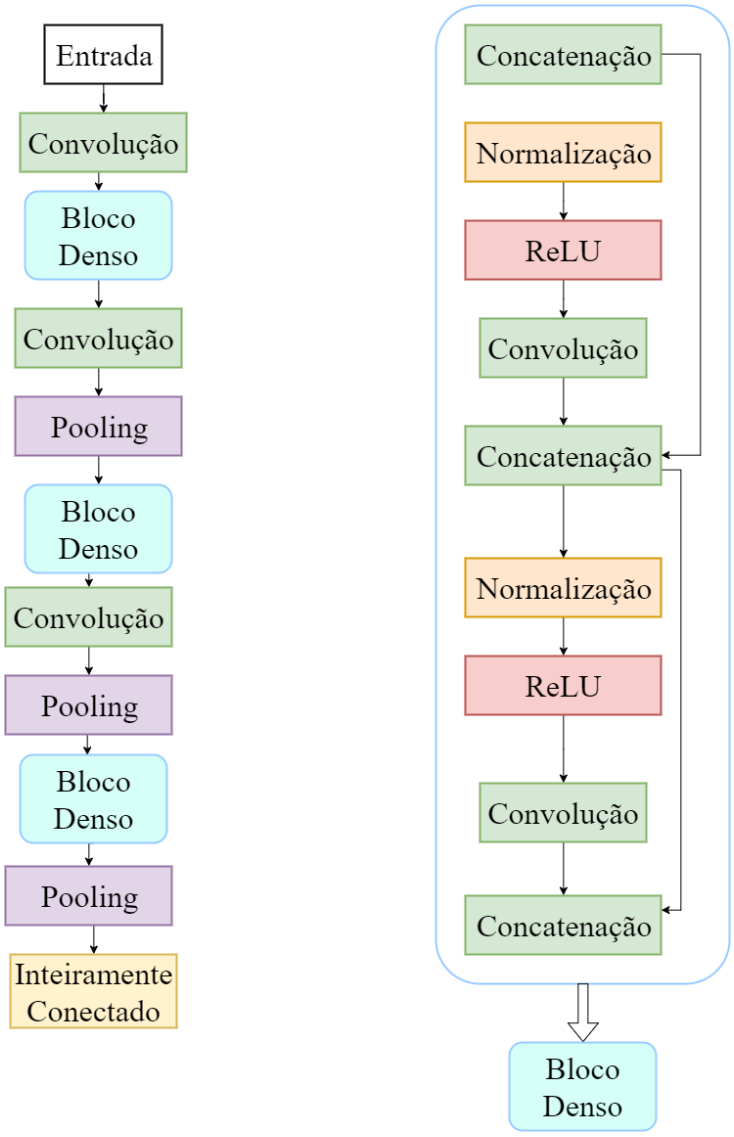
Fonte: Adaptado de (HE et al., 2015).

2.3.5 DenseNet

O modelo desenvolvido por (Huang et al., 2016) tem como inspiração a ResNet. Esse modelo utiliza o mesmo princípio de preservar a informação para as camadas mais profundas. Em vez de utilizar a adição para preservar a informação do bloco anterior, a saída desse bloco, chamado bloco denso, é concatenada com o resultado, como pode-

mos ver na Figura 2.12. Para que a informação possa ser enviada para camadas com diferentes tamanhos é utilizada uma camada de transição, que aplica filtros com tamanho 1×1 seguido por *average pooling*. Isso permite um melhor fluxo de informação pela rede, tornando a rede mais fácil de treinar.

Figura 2.12 – Arquitetura da DenseNet.



Fonte: Adaptado de (Huang et al., 2016).

2.3.6 SqueezeNet

O trabalho realizado por (IANDOLA et al., 2016) tem como objetivo desenvolver uma rede neural com a mesma acurácia da AlexNet, mas com 50 vezes menos parâmetros e

que o modelo treinado possa ser salvo utilizando menos de 1 Megabyte. As vantagens que esse modelo apresenta são:

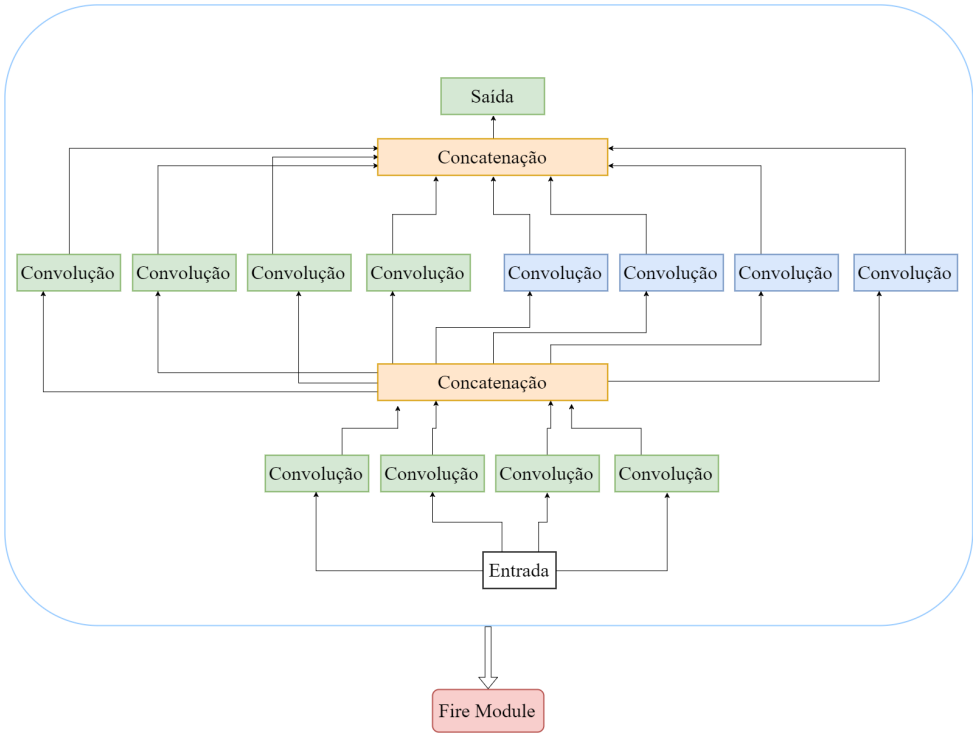
- Maior eficiência ao treinar a rede utilizando múltiplas unidades de processamento gráfico em paralelo. Uma vez que a quantidade de parâmetros a ser transferida entre essas unidades é menor.
- Facilidade ao transferir o modelo já treinado via internet. Possibilitando atualizações mais frequentes a sistemas embarcados que utilizam esse modelo.
- Funcionar em sistemas que não possuem muita memória disponível, como sistemas embarcados.

Para tornar isso possível foram empregadas algumas técnicas como:

- Utilizar muitos filtros com tamanho 1×1 em vez de filtros com tamanho 3×3, uma vez que eles possuem 9 vezes menos parâmetros.
- Reduzir a quantidade de canais de entrada que filtros com tamanho 3×3 recebem.
- Aplicar camada de *pooling* mais para o final do modelo.

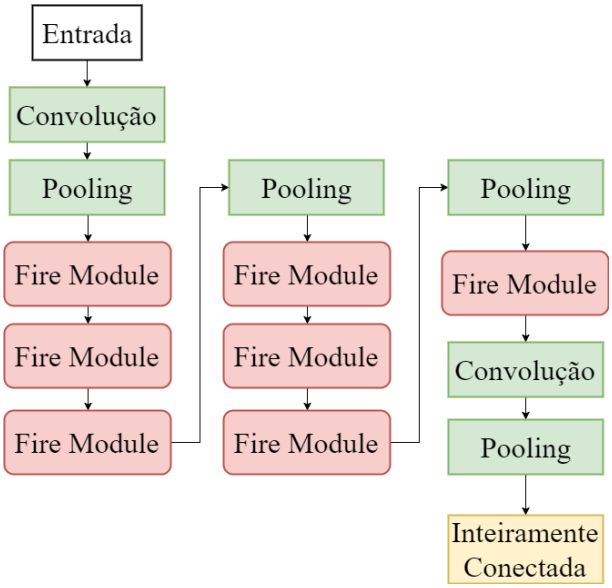
A principal característica dessa rede é um bloco de camadas chamado *Fire Module*, como podemos ver na Figura 2.13. Esse bloco é composto por dois sub-blocos: *Squeeze Module* e *Expand Module*. O primeiro é composto por uma camada de filtros com tamanho 1×1, responsável por reduzir o número de canais que será passado para o próximo bloco. O segundo bloco é composto por filtros de tamanho 1×1, em verde e 3×3, em azul, aplicados a saída do bloco anterior e esse resultado então é concatenado. A arquitetura da rede é principalmente formada pelos *Fire Module*, como podemos ver na Figura 2.14

Figura 2.13 – Fire Module.



Fonte: Adaptado de (IANDOLA et al., 2016).

Figura 2.14 – SqueezeNet.



Fonte: Adaptado de (IANDOLA et al., 2016).

2.4 DETECÇÃO DE OBJETOS

Imagens digitais nem sempre apresentam exclusivamente o objeto que desejamos utilizar para realizar algum procedimento. Muitas vezes o objeto que buscamos representa uma pequena porção da imagem e está junto de muitos outros objetos. Isso pode ser um problema ao aplicarmos algum algoritmo de classificação de objetos, devido a grande quantidade de ruído gerada por outras áreas da imagem. Para que esse tipo de problema pudesse ser evitado foram desenvolvidos métodos de detecção de objetos em imagens. Esses métodos visam buscar o objeto desejado para que seja possível aplicar algum algoritmo de forma isolada, nesse objeto. Esses métodos também permitiram com que vários objetos pudessem ser detectados em uma mesma imagem, possibilitando sistemas avançados como o de carros autônomos.

Um dos meios mais simples de construirmos um sistema de detecção de objetos é através de janelas deslizantes. Esse método desliza uma região, com tamanho fixo ou variado, por toda a imagem, dividindo-a em micro-regiões. Podem ser utilizadas técnicas de redimensionamento para aumentar a chance de encontrar o objeto independente de seu tamanho. Cada micro-região é processada por um sistema de classificação de objetos, caso o sistema encontre o objeto as coordenadas dessa região são salvas. Apesar desse sistema funcionar ele contém alguns problemas, principalmente de performance, esse método aplicado em uma imagem com dimensões 200×200, com uma janela deslizante com dimensões 10×10, deslizando essa janela metade de suas dimensões por avanço, aplica o classificador 1600 vezes, não considerando redimensionamentos. Cada aplicação do classificador é um processo custoso, dificultando o uso desse método em aplicações que demandam a detecção em tempo real.

2.4.1 YOLO9000

Devido a grande demanda por sistemas de detecção de objetos, surgiram métodos alternativos para contornar os problemas do método anterior.

O método desenvolvido por (Girshick et al., 2013) chamado de *Region-based Convolutional Neural Network* (R-CNN) foi um dos primeiros métodos a utilizar RNC para esse propósito. Esse método utiliza um algoritmo de busca seletiva, agrupando regiões com pixeis similares, para reduzir o número de regiões em que o classificador deve atuar, e um algoritmo simples de aprendizado de máquina para classificar essas regiões.

O trabalho de (Redmon et al., 2015), chamado de *You Only Look Once* (YOLO), demonstra uma alternativa ao método anterior, com o foco no desempenho de detecção. Uma única RNC é responsável por localizar todos os objetos da imagem, não sendo necessário treinar outro algoritmo de aprendizado de máquina igual ao R-CNN. Esse método

processa a toda a imagem, em vez de partes previamente selecionadas, extraindo informações contextuais sobre os objetos, melhorando a distinção entre o fundo da imagem e o objeto. O algoritmo divide a imagem de entrada em uma grade de tamanho A$\times$A, onde o valor padrão de A é 7. Se o grande parte do objeto estiver em um dos blocos da grade então esse bloco detectará esse objeto. Cada bloco utiliza B caixas delimitadoras com tamanhos diferentes, onde B tem valor padrão de 2. Esses blocos pontuam a região em 5 valores: coordenadas x e y, altura e largura da região e probabilidade de existir um objeto. Caso haja a presença de um objeto, são realizados ajustes nas dimensões do bloco, para se aproximar do valor anotado.

O método desenvolvido por (REDMON; FARHADI, 2016), chamado de YOLO9000, apresenta uma versão aprimorada do método anterior. Uma técnica de normalização foi utilizada na rede aumentando a acurácia. A rede foi treinada em um grande banco de imagens com dimensões 224×224, para melhorar o classificador. Então foi retreinada em um banco de imagens com dimensões 448×448 para realizar um ajuste fino, melhorando a precisão em imagens com dimensões maiores. Em vez do sistema de grade é utilizado o sistema de âncoras. As camadas inteiramente conectadas foram removidas, a rede é redimensionada para ter dimensões 416×416, pois o processamento das camadas intermediarias reduz em 32 vezes essas dimensões para 13×13. Por ter dimensões impares possui um centro, facilitando a detecção de objetos, que geralmente estão no centro da imagem. Em cada uma das áreas criadas é marcado um centro, em cada centro 5 caixas com tamanhos distintos são criadas, que verificará a presença de um objeto na área. O ajuste preciso do tamanho dessas caixas é feito por um algoritmo de agrupamento. A remoção da camada completamente conectada possibilita com que a rede possa ser redimensionada durante o treinamento, melhorando a acurácia em imagens de diferentes dimensões.

2.5 CONSIDERAÇÕES

A área de visão computacional teve grandes avanços na última década. Avanços que em grande parte foram obtidos da área de aprendizado de máquina.

Para que um objeto possa ser classificado com a maior acurácia possível é necessário que o sistema tenha sido treinado em grandes quantidades de imagens. Apesar da quantidade disponível de imagens digitais, rotuladas por especialistas, ter aumentado drasticamente nos últimos tempos, ainda temos dificuldade para encontrar grandes bases de dados públicas, principalmente da área da medicina. Então por enquanto esse problema só pode ser amenizado utilizando arquiteturas mais robustas.

Também é possível utilizar um conjunto de técnicas para diminuir o ruido do banco de imagens. Como a detecção de objeto para detectar apenas uma região da imagem,

chamada de RI, facilitando o trabalho do classificador.

A métrica de curva COR foi escolhida após a análise de trabalhos relacionados, uma vez que grande parte deles adotou essa métrica para a análise da performance de seus classificadores. A matriz de confusão foi escolhida por apresentar uma visão geral dos acertos e erros do classificador, de uma maneira fácil de se visualizar.

O Capítulo 3, apresenta trabalhos relacionados sobre sistemas de detecção de objetos e sistemas classificadores de objetos.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta na seção 3.1 a descrição das métricas utilizadas pelos trabalhos relacionados, assim como, os trabalhos existentes que realizam a detecção da Região de Interesse (RI), vistos na seção 3.2, e a detecção do glaucoma na seção 3.3.

3.1 MÉTRICAS DE AVALIAÇÃO

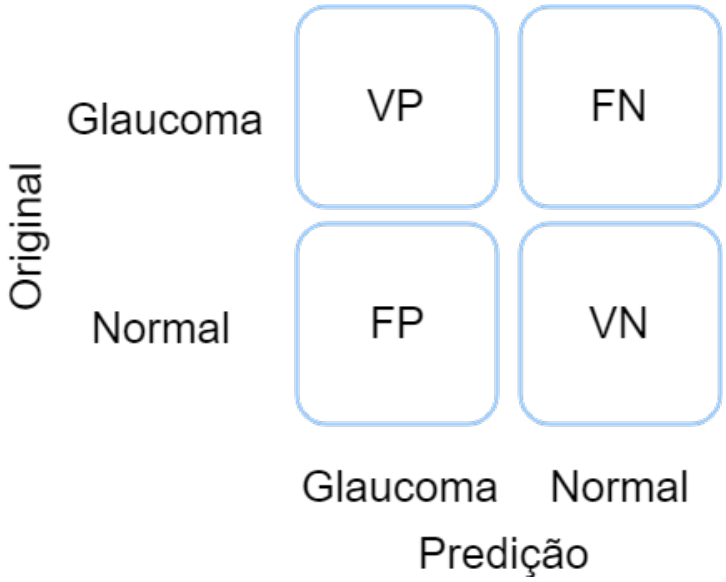
As métricas de avaliações permitem que possamos analisar pouco mais a fundo o como o classificador está se comportando em relação aos dados processados. Muitas vezes apenas contar a quantidade de acertos de uma classe não representa a verdadeira performance do classificador, pois isso depende da distribuição entre as amostras.

3.1.1 Matriz de confusão

A matriz de confusão é uma tabela que é geralmente utilizada para descrever a performance de um classificador em um conjunto de dados, cujo os rótulos de cada classe foram previamente anotados. Essa tabela apresenta um sumário dos resultados de predição do classificador, como pode ser visto na Figura 3.1. A seguir os termos utilizados para a construção da matriz de confusão:

- Verdadeiro Positivo (VP), são os casos em que a imagem com glaucoma foi detectada corretamente com glaucoma.
- Falso Positivo (FP), a imagem com glaucoma foi detectada como normal.
- Verdadeiro Negativo (VN), a imagem normal foi detectada como normal.
- Falso Negativo (FN), a imagagem normal foi detectada com a presença de glaucoma.

Figura 3.1 – Matriz de confusão.



Fonte: Autoria própria.

A partir dessa tabela podemos extrair algumas métricas como:

- Acurácia, dada pela Equação 3.1, demonstra a quantidade de amostras classificadas corretamente pelo classificador.
- Sensibilidade, dada pela Equação 3.2, mede a quantidade de imagens com glaucoma, VP, que foram corretamente identificadas, em relação ao total de imagens com glaucoma.
- Especificidade, dada pela Equação 3.2, mede a quantidade de imagens normais, VN, que foram corretamente identificadas, em relação ao total de imagens normais.

$$\frac{VP + VN}{VP + VN + FP + FN} \tag{3.1}$$

$$\frac{VP}{VP + FN} \tag{3.2}$$

$$\frac{VN}{VN + FP} \tag{3.3}$$

3.1.2 Curva de Característica de Operação do Receptor (COR)

Segundo (CERDA; CIFUENTES, 2012) o gráfico da curva COR demonstra a sensibilidade, no eixo Y e 1 menos a especificidade, no eixo X. Esse gráfico apresenta cada um dos possíveis pontos de corte de um teste de diagnóstico, cuja escala de medição é contínua. A linha diagonal, que vai do ponto 0.0, 0.0 até o ponto 1.0, 1.0 indica uma classificação aleatória, não confiável. Para que a acurácia da RNC seja considerado boa a curva precisará estar acima dessa linha. A área abaixo dessa curva mede a performance de discriminação da RNC e normalmente é utilizada como medida padrão para comparar classificadores. Valores entre 0.9 e 1.0 são considerados excelentes, entre 0.8 e 0.9 como bons, entre 0.7 e 0.8 como razoável, entre 0.6 e 0.7 como ruins e entre 0.5 e 0.6 como falhou.

3.2 DETECÇÃO DA REGIÃO DE INTERESSE

O trabalho realizado por (SINTHANAYOTHIN et al., 1999) tem como objetivo identificar automaticamente os principais componentes de uma imagem do fundo do olho. Um desses componentes é o disco óptico. Foram utilizadas 112 imagens da retina com dimensões de 570×550 pixeis com três canais de cores, RGB (Red Green Blue). Essas imagens foram preprocessadas utilizando a técnica de contraste adaptativo local para normalizar a intensidade da imagem. O disco óptico foi localizado identificando a área com maior variação de intensidade de pixeis adjacentes. O disco óptico foi encontrado com 99,1% de acurácia.

O método de (WALTER; KLEIN, 2001) apresenta um algoritmo baseado em morfologia matemática para a detecção do disco óptico e dos vasos sanguíneos em imagens ruidosas e de baixo contraste. Foi utilizado o canal vermelho de imagens RGB, pois ele apresenta um pequeno alcance dinâmico e o disco óptico geralmente pertencer a parte com mais brilho da imagem. É utilizado um filtro de binarização na imagem para obter o disco óptico assim como outras lesões na imagem. É afirmado que essas lesões não causam problemas na detecção, pois são muito menores que o disco óptico. São utiliza-

dos filtros na região do disco para preencher os vasos sanguíneos, tornado a região mais uniforme. Por fim é aplicado o algoritmo clássico de segmentação de imagens, chamado *watershed*, para achar os contornos do disco óptico. O método foi testado em 30 imagens coloridas com dimensões de 640×480 pixeis contendo várias patologias. Em 27 imagens foi encontrado o contorno exato e em 3 houveram erros devido ao baixo ou alto contraste do canal vermelho.

A técnica utilizada por (LALONDE; BEAULIEU; GAGNON, 2001) tem o propósito de identificar o disco óptico em imagens do fundo do olho de baixa resolução. As imagens foram convertidas para escala de cinza utilizando o canal verde como referência. As dimensões da imagem foram reduzidas para que as características importantes fossem concentradas. Os pixeis com maior intensidade são comparados com a intensidade média da imagem para encontrar essas regiões. Para cada região proposta foi utilizado um método de segmentação para achar os contornos. Então foi aplicado um algoritmo de busca de padrões utilizando à distância de Hausdorff para verificar se a área correspondia ao disco óptico. O método foi testado em 40 imagens do fundo do olho coloridas com diferentes qualidades. Foi encontrado um erro de 7% em relação ao posicionamento central do disco, sem falsas detecções.

(CHAI et al., 2017) utiliza um método baseado em rede neural para extrair o a RI utilizada em seu trabalho sobre a detecção do glaucoma. Para utilizar esse método 1500 imagens do fundo do olho tiveram sua RI anotada manualmente pelos autores. Então a rede neural *Faster-RCNN*, que utiliza um princípio similar ao da YOLO apresentada anteriormente, é treinada utilizando 1200 dessas imagens e testada em 300. Não mais do que 10 imagens falharam em ter sua RI detectada.

O trabalho de (ORLANDO et al., 2017) utiliza uma técnica em janela deslizante para a detecção da RI. Essa técnica consiste em deslizar uma região de tamanho fixo sobre a imagem, variando a escala dessa imagem. A região percorre a imagem avançando em passos com um distância predefinida. Para cada passo uma sub-região é gerada, esta região então é classificada utilizando a RNC Inception v3. Cada imagem utilizada no treinamento foi recortadas manualmente, separando a RI de vários recortes do restante da imagem do fundo do olho. Esse processo gerou 107 imagens da ROI e 4693 de outras regiões, utilizando todas as imagens do banco HRF. A rede atingiu 99% de acurácia de classificação. Para evitar que a mesma RI fosse detectada mais de uma vez, foi utilizado um método de supressão que tem como objetivo demarcar a área média da RI detectada múltiplas vezes.

3.3 DETECÇÃO DO GLAUCOMA

(ORLANDO et al., 2017) faz um estudo sobre duas RNC diferentes, pre-treinadas em dados não médicos, para a detecção do glaucoma. A primeira rede se chama OverFeat e possui 6 camadas convolucionais e utiliza técnicas de *max pooling*. A segunda com nome de VGG-S possui arquitetura similar mas com 5 camadas convolucionais e menos filtros em suas últimas camadas. As redes foram pre-treinadas utilizando o banco de imagens *ImageNet* 2012, que consiste em 1,2 milhão de imagens divididas em 1000 diferentes classes. Esse processo tem como objetivo reduzir o tempo de treinamento e a quantidade de imagens necessárias, utilizando a técnica de transferência de aprendizado. As imagens foram pré-processadas utilizando equalização adaptativa de histograma com limitação de contraste e remoção dos vasos sanguíneos. Também foram utilizadas técnicas de aumento de dados, como rotação e espelhamento, aumentando o tamanho do banco de imagens de 8 a 16 vezes. O banco de imagens Drishti-GS1, com 101 imagens com dimensões de 2896×1944 pixeis, coletado do Aravind Eye Hospital na Índia, foi utilizado para a validação do processo. Utilizando 30% das imagens para teste e 70% para o treinamento a OverFeat conseguiu 76,265% e a VGG-S 71,8% de acurácia no melhor caso.

O trabalho de (CHAI et al., 2017) utiliza uma RNC com duas ramificações para detectar o glaucoma. Cada uma é um modelo clássico de RNC formada por 5 camadas convolucionais, assim como *max pooling* e ativações ReLU. Uma ramificação recebe apenas a RI, extraída por outra rede neural. A outra ramificação recebe a imagem inteira do fundo do olho. As duas ramificações são concatenadas, combinando as características extraídas de cada uma. O banco de imagens utilizado contém 3554 imagens do fundo do olho, com tamanhos variados, de aproximadamente 2000 pacientes que sofrem de doenças variadas como: glaucoma, catarata e retinopatia diabética. As imagens utilizadas no experimento foram coletadas do *Beijing Tongren Hospital* (BTH). Dessas imagens 1391 foram diagnosticadas com glaucoma. Para o treinamento a RI foi redimensionada para 128×128 pixeis, enquanto as imagens do fundo do olho foram redimensionadas para 256×256 pixeis. Utilizando 20% das imagens para teste e 80% para treino a acurácia encontrada foi de 81.69%.

(CHEN et al., 2015) propõe uma arquitetura de RNC para a detecção do glaucoma. A arquitetura é composta de 4 camadas convolucionais e duas inteiramente conectadas. Técnicas de aumento de dados também foram utilizadas, como: cinco recortes fixos de cada imagem e espelhamento horizontal. Dois bancos de imagens foram utilizados, ORIGA com 168 imagens com glaucoma e 482 normais, e SCES com 1676 imagens normais e 46 com glaucoma. A acurácia dos experimentos nos bancos de imagem foi de 83,1% com o ORIGA e 88,7% com o SCES.

O método descrito por (SHEEBA et al., 2014) utiliza uma rede neural simples de duas camadas para a detecção. São utilizados parâmetros extraídos manualmente de

imagens como: pressão intraocular, espessura da córnea, espessura da fibra do nervo e razão entre o tamanho da escavação com o tamanho do disco. A rede foi treinada utilizando parâmetros de 20 pacientes. A validação ocorreu em 28 imagens de pacientes com glaucoma e 12 normais, das quais 34 dessas imagens foram classificadas corretamente, obtendo 85% de acurácia. Essas imagens foram retiradas do *Giridhar Eye Institute* localizado em Cochim, Índia.

(DUTTA et al., 2014) apresenta um método baseado em segmentação para a detecção do glaucoma. Para a segmentação do disco óptico e da escavação foram utilizados dois limiares: um para remover os vasos sanguíneos e outras partes não relevantes, o segundo para segmentar os pixeis de alta intensidade da RI, separando a região da escavação da região do disco óptico. Então a Transformada de Hough foi utilizada para encontrar o raio de cada região. Imagens com o raio da escavação dividido pelo raio do disco óptico com resultados superiores a 0,75 foram consideradas com a presença do glaucoma. O banco de imagens High Resolution Fundus (HRF) foi utilizado para a validação do algoritmo, atingindo 90% de acurácia.

O trabalho de (ALLAN et al., 2017) utiliza a RNC Inception v3 para a detecção do glaucoma. O experimento foi realizado utilizando os bancos de imagem: HRF, RIM-ONE R1, RIM-ONE R2 e RIM-ONE R3, mais detalhes sobre esses bancos de imagens são apresentados na seção 4.1. A classificação é realizada através RI, que por sua vez é extraída por um método de janela deslizante. Nos bancos de imagens com menos de 100 imagens foi utilizado um algoritmo de aumento de dados, que aplica distorções como: espelhamento, recortes aleatórios, variação do gama, adição de ruído Gaussiano e rotações de 90º. A rede foi treinada utilizando 90% das imagens de cada banco e validada com o restante. A acurácia foi de 90% para o banco HRF, 94,2% para o RIM-ONE R1, 86,2% para o RIM-ONE R2, 86,4% para o RIM-ONE R3 e 87,6% para a fusão de todos os bancos.

3.4 CONSIDERAÇÕES

Em relação à detecção do disco óptico podemos notar na Tabela 3.1 que o método mais comum é o de localizar a maior região de pixeis com maior intensidade. Esse método geralmente funciona bem em imagens de boa qualidade e apresentando poucas lesões. Existem lesões que atingem grande parte do fundo do olho, sendo algumas vezes maior do que o disco óptico, tornando esse método sensível a imagens desse tipo. O trabalho de (ALLAN et al., 2017) utiliza o método de janela deslizante, apesar desse método atingir uma boa acurácia ele possui grandes problemas de performance ao utilizar a RNC Inception v3 para detectar a RI. Para cada região proposta o classificador realiza a detecção, para o método obter uma maior precisão é necessário diminuir o tamanho de cada avanço aumentando ainda mais a quantidade de regiões proposta. O trabalho de (CHAI

et al., 2017) utiliza a rede Faster-RCNN para a detecção da RI, esse método não apresenta os problemas dos anteriores mas existem versões mais eficientes para a detecção de objetos. Esse trabalho utiliza a rede YOLO9000, que segundo os experimentos realizados por (REDMON; FARHADI, 2016) possui mais performance do que a Faster-RCNN mantendo acurácia equivalente, isso possibilita com que futuros trabalhos possam implementar o método da detecção de RI em sistemas com poucos recursos computacionais ou embarcados.

No que diz respeito a detecção do glaucoma a Tabela 3.2 demonstra que os trabalhos recentes têm utilizado RNC para este fim. Métodos baseados em segmentação não são muito estáveis, devido ao processo de segmentação ser sensível a ruídos e artefatos nas imagens. O banco de imagens HRF apresenta imagens sem esses problemas, então métodos como o de (DUTTA et al., 2014) podem não funcionar em outros bancos de imagens sem alguns ajustes. Muitos dos trabalhos utilizando RNC também utilizam bancos de imagens que o autor não conseguiu acesso, dificultando a comparação entre os resultados. O ideal seria que os experimentos realizados por esses autores utilizassem bancos de imagens públicos. Esse trabalho realiza uma avaliação de acurácia entre diferentes RNC em bancos de imagens públicos, afim de encontrar a com mais precisão para realizar o diagnóstico e a com melhor performance enquanto mantém boa acurácia. Isso possibilita com que trabalhos futuros utilizem a RNC com o melhor custo benefício para possíveis sistemas implementando a detecção do glaucoma.

Tabela 3.1 – Comparativo dos métodos utilizados para encontrar o disco óptico.

Autores	**Método**	**Acurácia**
(SINTHANAYOTHIN et al., 1999)	Baseado na intensidade dos pixeis.	99,10%
(WALTER; KLEIN, 2001)	Baseado na intensidade dos pixeis.	90,00%
(LALONDE; BEAULIEU; GAGNON, 2001)	Baseado na intensidade dos pixeis.	100%
(CHAI et al., 2017)	RNC.	96,60%
(ALLAN et al., 2017)	Janela deslizante.	99,00%

Fonte: Autoria própria.

Tabela 3.2 – Comparativo dos métodos utilizados para detecção do glaucoma.

Autores	**Método**	**Acurácia**	**Banco de Imagem**
(ORLANDO et al., 2017)	RNC OverFeat.	76,26%	Drishti-GS1
(ORLANDO et al., 2017)	RNC VGG-S.	71,80%	Drishti-GS1
(CHAI et al., 2017)	RNC com duas ramificações.	81,69%	BTH
(CHEN et al., 2015)	RNC customizada.	88,70%	SCES
(CHEN et al., 2015)	RNC customizada.	83,10%	ORIGA
(SHEEBA et al., 2014)	RNA com características anotadas.	85,00%	Giridhar Eye Institute
(DUTTA et al., 2014)	Segmentação.	90,00%	HRF
(ALLAN et al., 2017)	RNC Inception v3.	90,00%	HRF
(ALLAN et al., 2017)	RNC Inception v3.	94,20%	RIM-ONE-R1
(ALLAN et al., 2017)	RNC Inception v3.	86,20%	RIM-ONE-R2
(ALLAN et al., 2017)	RNC Inception v3.	86,40%	RIM-ONE-R3
(ALLAN et al., 2017)	RNC Inception v3.	87,60%	Todos

Fonte: Autoria própria.

4 MATERIAIS E MÉTODOS

Este capítulo apresenta os bancos de imagens utilizados, descritos na seção 4.1, para os experimentos. O equipamento utilizado é exibido na seção 4.2. Na seção 4.3 são apresentadas as ferramentas utilizados os experimentos desse trabalho. As seções 4.4 e 4.5 apresentam os métodos utilizados na detecção da RI e do glaucoma, respectivamente.

4.1 BANCOS DE IMAGENS

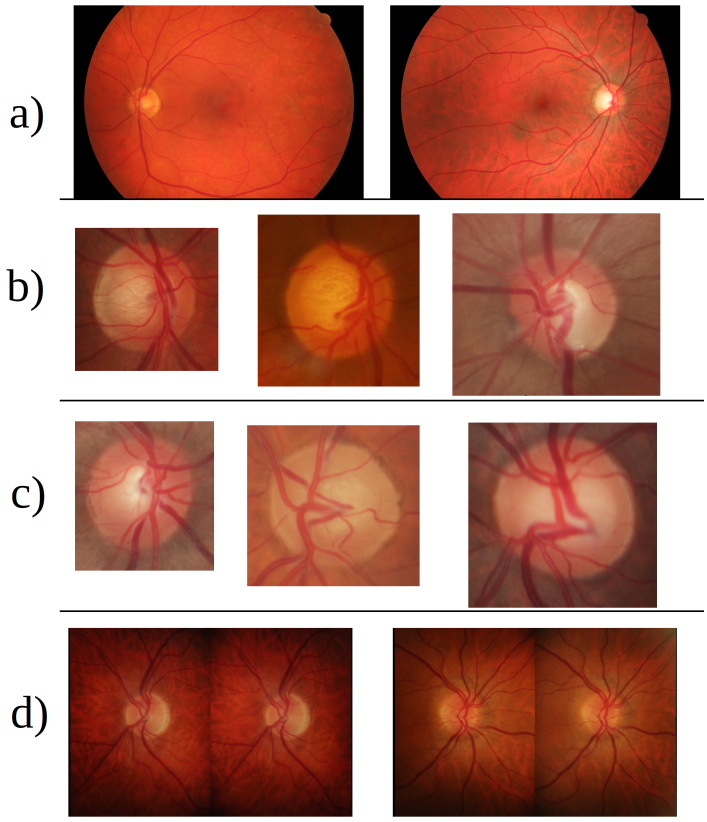
O trabalho visava utilizar os mesmos bancos utilizados pelos trabalhos relacionados para a avaliação da acurácia, mas como grande parte desses bancos são privados ou o autor não conseguiu entrar em contato para solicitar os mesmos, o trabalho buscou por bancos de imagens públicas. Isso possibilita com que trabalhos futuros possam comparar seus resultados com os do trabalho atual.

O primeiro banco utilizado se chama *High Resolution Backgrounds* (HRF) e foi elaborado por (KOLAR et al., 2013). Esse banco é composto por 45 imagens do fundo do olho divididas em 3 grupos de 15 imagens. Um grupo contém imagens diagnosticadas com glaucoma, outro com retinopatia diabética e o último apresenta imagens saudáveis.

O trabalho de (FUMERO et al., 2011) apresenta 3 bancos de imagens chamados RIM-ONE R1, RIM-ONE R2 e RIM-ONE R3. Os primeiros dois bancos apresentam imagens da RI e o terceiro apresenta imagens em estéreo do olho diagnosticado do paciente. RIM-ONE R1 é composto por 40 imagens com algum grau de glaucoma e 118 normais. RIM-ONE R2 é composto por 200 imagens com glaucoma e 225 normais. RIM-ONE R3 é composto por 74 imagens com glaucoma e 85 normais.

A Figura 4.1 apresenta alguns exemplos de imagens encontradas nos bancos apresentados.

Figura 4.1 – Exemplos de imagens encontradas nos bancos utilizados nesse trabalho. a) HRF. b) RIM-ONE R1. c) RIM-ONE R2. d) RIM-ONE R3.



Fonte: Autoria própria.

4.2 EQUIPAMENTO UTILIZADO

Para os experimentos realizados nesse trabalho foi utilizado um computador com as seguintes características:

- Processador: AMD FX-8350 com 8 núcleos físicos rodando a 4,2 GHz com 8MB de cache L2.
- Memória: 16 GB DDR3 1600 MHz.
- Placa de Vídeo: Nvidia GeForce GTX 1070.
- Armazenamento: SSD Samsung 850 EVO com 250GB.
- Sistema Operacional: Ubuntu 16.04.3 LTS.

4.3 FERRAMENTAS

Essa versão utiliza o *framework* de código aberto para aprendizado de máquina chamado de PyTorch (PASZKE et al., 2017), desenvolvido pela empresa Facebook. Esse foi originalmente desenvolvido, com o nome de Torch, utilizando a linguagem de programação Lua e foi reescrito para Python com o nome de PyTorch. Esse *framework* foi escolhido pois apresentava suporte a uma versão da API CUDA, responsável por realizar cálculos paralelos utilizando a placa de vídeo, que ainda não era suportada por outros frameworks como o Tensorflow. O PyTorch também foi escolhido por apresentar uma grande variedade de algoritmos e métodos do estado da arte implementado pela comunidade.

Para o aumento de dados foi utilizado uma biblioteca de código aberto desenvolvida em Python com nome de Augmentor (BLOICE; STOCKER; HOLZINGER, 2017). Essa biblioteca foi desenvolvida facilitar o processo de aumento de dados para sistemas de aprendizado de máquina que processam imagens.

Para a detecção da RI foi utilizado um *framework* de código aberto chamado de Darknet, desenvolvido pelo mesmo autor da rede YOLO9000, (REDMON; FARHADI, 2016), utilizando a linguagem de programação C e CUDA.

4.4 DETECÇÃO DA RI

Para que objetos possam ser detectados utilizando a rede YOLO9000 é preciso que essa rede seja previamente treinada. O processo de treinamento dessa rede ocorre de maneira diferente das redes convencionais. A mesma precisa ter conhecimento da localização de um ou mais objetos na imagem. Essa localização é demarcada por 5 parâmetros numéricos: número da classe do objeto, coordenada X e Y do centro do objeto, proporção entre a largura e a altura do objeto em relação as dimensões da imagem.

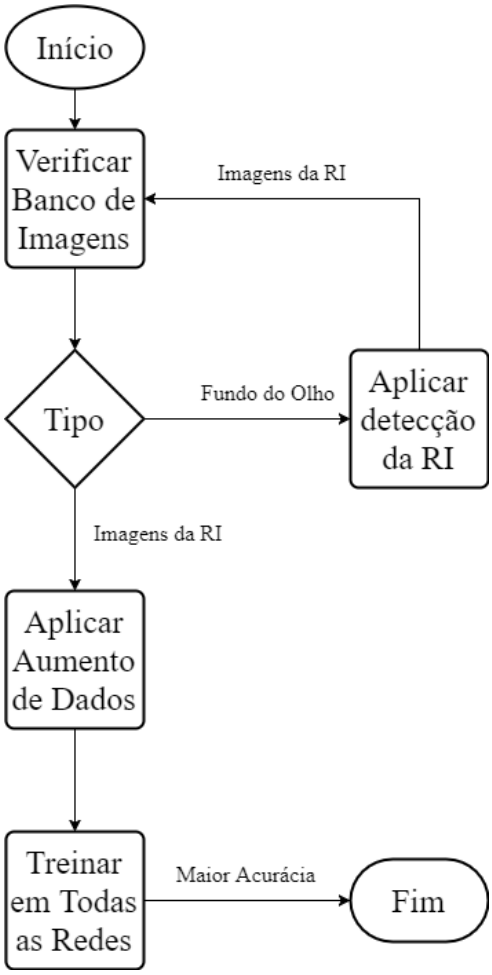
O banco de imagens RIM-ONE r3 foi escolhido para o treinamento, pois apresenta imagens com grande variação de luminosidade e formatos diferenciados da RI. As coordenadas de todas as imagens foram extraídas manualmente. A rede foi treinada utilizando os parâmetros padrões. Após o treinamento a rede foi capaz de detectar a RI de todas as imagens do banco HRF.

4.5 DETECÇÃO DO GLAUCOMA

Para a detecção do glaucoma cada banco de imagens seguiu o fluxo apresentado na Figura 4.2. As redes precisam da RI como imagem de entrada, caso o banco apresente

imagens do fundo do olho o algoritmo de detecção da RI é utilizado. A técnica de aumento de dados foi utilizada em todos os bancos, aumentando a quantidade de imagens em 5 vezes mais do que a original. Cada banco de imagens foi treinado em todos as arquiteturas apresentadas na seção 2.3. O PyTorch possui todas essas arquiteturas já implementadas, facilitando o processo de treinamento. O banco de imagens RIM-ONE r1 foi agrupado com o RIM-ONE r2 pois ambos apresentam imagens da RI e grande similaridade, como pode ser visto na Figura 4.1.

Figura 4.2 – Diagrama do fluxo aplicado para o treinamento da versão final do trabalho.



Fonte: Autoria própria.

4.6 CONSIDERAÇÕES

Este capítulo apresentou os bancos de imagens, ferramentas e equipamentos utilizados para os experimentos deste trabalho.

O próximo capítulo apresenta os resultados obtidos por cada RNC em cada banco de imagens, assim como a comparação dos resultados.

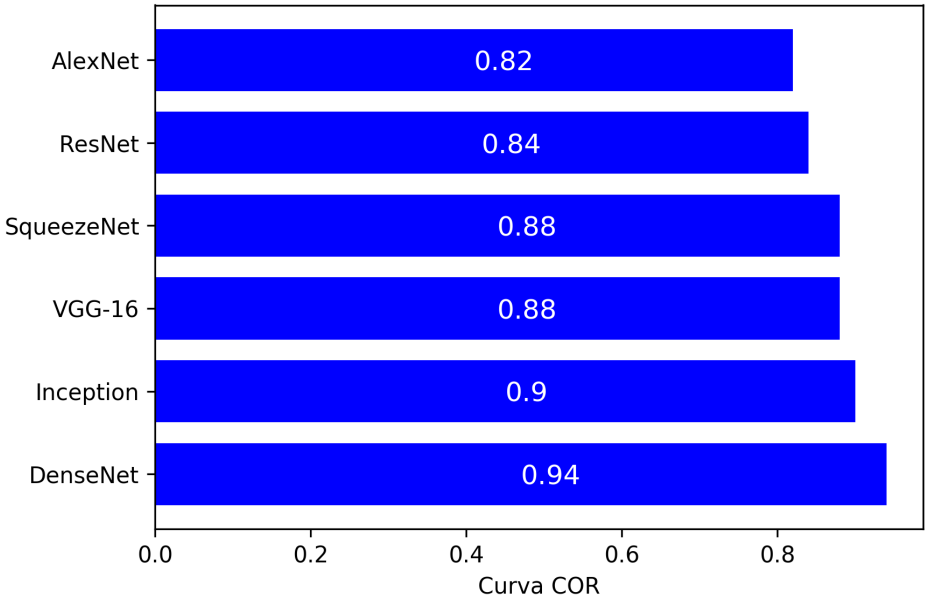
5 RESULTADOS

Este capítulo apresenta os resultados gerados pelos experimentos. Para a avaliação dos resultados foi utilizada a métrica mais comum dos trabalhos relacionados, a área sobre a curva COR. Ao final do capítulo são apresentadas as limitações do trabalho assim como a discussão dos resultados obtidos.

O processo de treinamento de uma RNA geralmente salva todos os pesos obtidos após a rede processar todas as imagens disponíveis para treinamento, esse processamento em todas as imagens é chamado de época. Após essa etapa geralmente a rede utiliza esses pesos salvos para classificar imagens ainda não vistas pela mesma, etapa chamadas de validação. Esse trabalho utiliza 90% das imagens para a etapa de treinamento e 10% para a etapa de validação. O resultado dessa etapa é comparado com o das etapas anteriores afim de encontrar a melhor combinação de pesos para a rede resolver o problema de classificação. Esse processo foi realizado 200 vezes em cada rede para cada banco de imagens.

Os resultados para a curva COR foram obtidos através de um *script* desenvolvido pelo autor, o mesmo verifica qual é a classificação da imagem original e compara com a classificação obtida pela RNC e gera o resultado. O resultado individual de cada experimento pode ser verificado no apêndice A. A Figura 5.1 apresenta uma comparação entre a média dos valores obtidos por cada rede, calculada pela soma das áreas da curva COR dividido pela quantidade de bancos de imagens avaliados.

Figura 5.1 – Comparação da média dos resultados das curvas ROC.



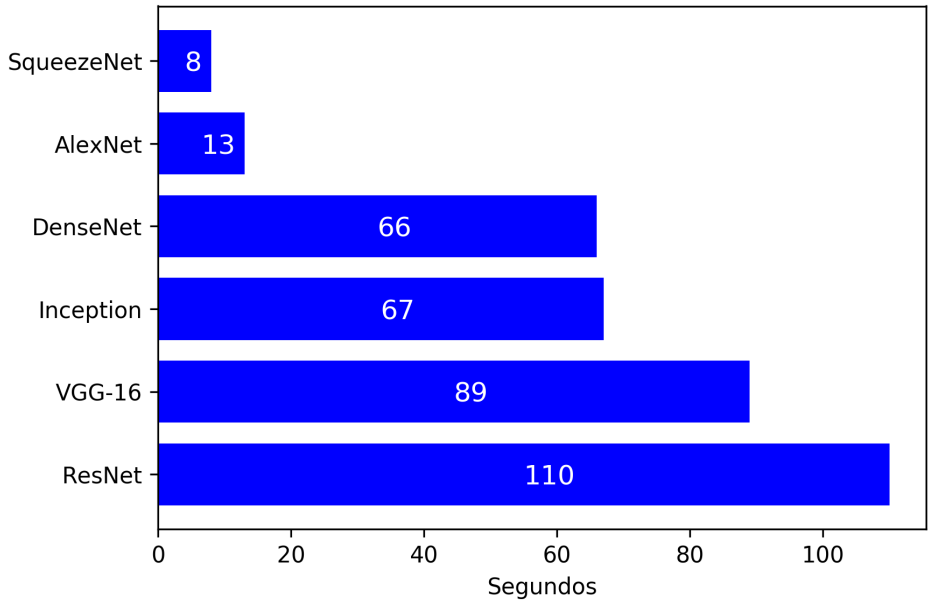
Fonte: Autoria própria.

5.1 OUTRAS COMPARAÇÕES

A seguir comparações de outros aspectos dos modelos de RNC. Os experimentos foram realizados utilizando as imagens de todos os bancos misturadas.

O gráfico da Figura 5.2 demonstra o tempo de treinamento de cada época. Uma época representa que a RNC observou todas as imagens disponíveis, reservadas para o treinamento, realizando os devidos ajustes. Cada época neste experimento é composta por 4740 imagens.

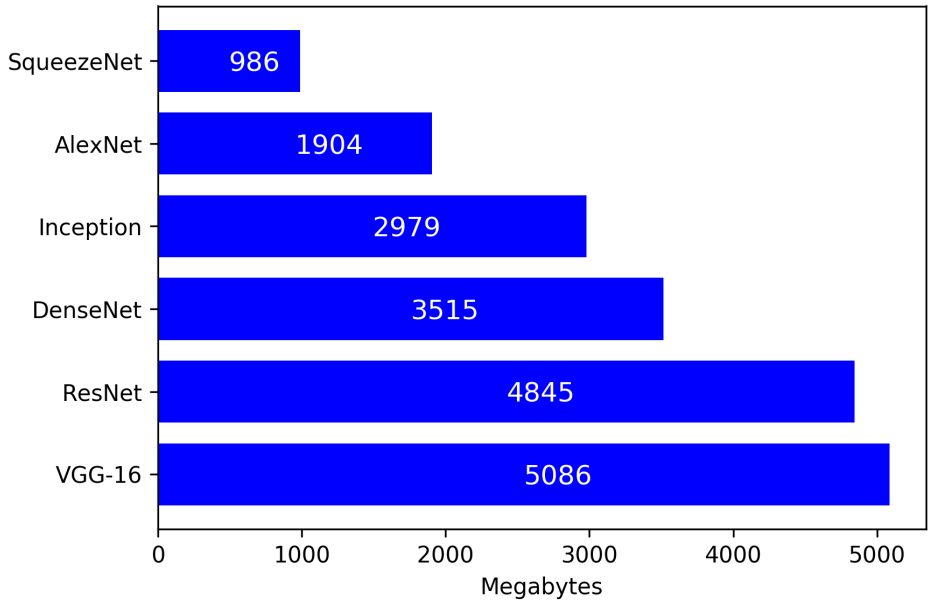
Figura 5.2 – Comparação do tempo de processamento de cada época durante o processo de treinamento.



Fonte: Autoria própria.

A Figura 5.3 apresenta a comparação do consumo de memória entre as RNC. O consumo foi medido através de um *software* que retorna informações da placa de vídeo. As medições foram executadas durante o processo de treinamento e a média aritmética foi utiliza para chegar ao valor. Esse valor não leva em consideração a memória ocupada por outras tarefas do sistema.

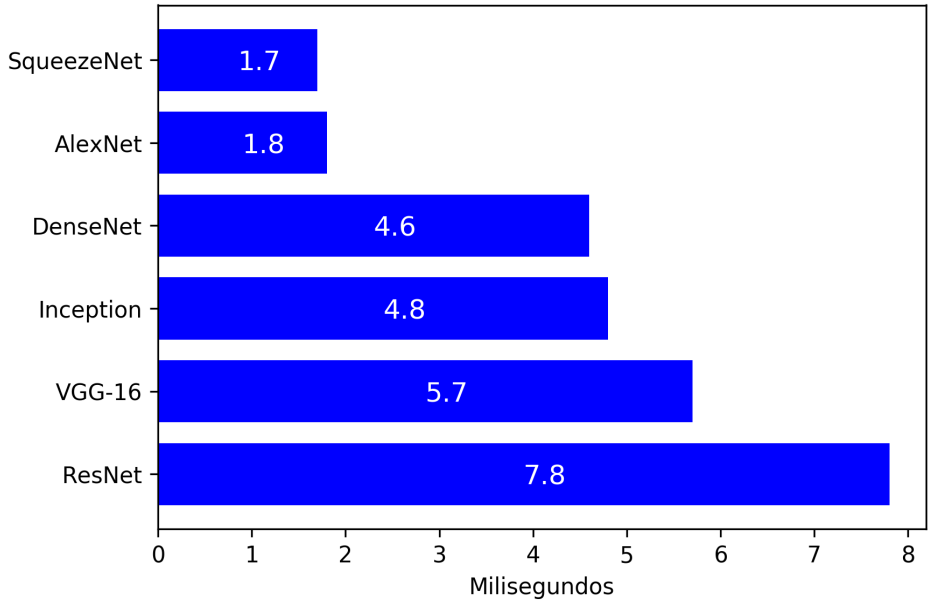
Figura 5.3 – Comparação do consumo de memória de cada modelo durante o processo de treinamento.



Fonte: Autoria própria.

O tempo médio para a RNC detectar a presença do glaucoma é apresentado na Figura 5.4. O experimento foi realizado utilizando o tempo para a RNC calcular a acurácia das 552 imagens de validação. Este tempo foi dividido pela quantidade de imagens para chegar no resultado apresentado.

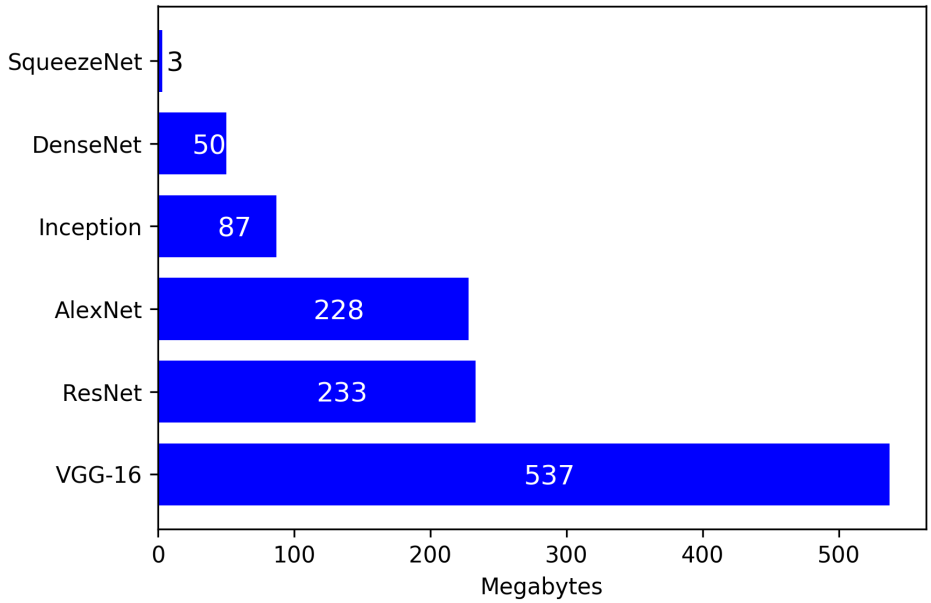
Figura 5.4 – Comparação do tempo médio de predição para uma imagem.



Fonte: Autoria própria.

A comparação do tamanho de cada modelo salvo em disco é apresentado na Figura 5.5. O arquivo salvo representa os pesos calculados pelo processo de treinamento, utilizados para realizar predições ou continuar o realizar o processo de ajuste fino na RNC.

Figura 5.5 – Comparação do tamanho ocupado em disco de cada modelo salvo.



Fonte: Autoria própria.

5.2 DISCUSSÕES

Este capítulo apresenta a análise dos dados coletados nos experimentos anteriores. Utilizando a média dos resultados encontrados na Figura 5.1 podemos perceber que a rede com maior acurácia de classificação geral é a DenseNet, seguida pela Inception e SqueezeNet empatada com a VGG-16. A DenseNet é a RNC mais nova de todas as utilizadas. Ela possui uma arquitetura inspirada na ResNet aplicando técnicas do estado da arte para garantir sua robustez. Utilizando os dados obtidos no apêndice A sobre a DenseNet, podemos comprar os resultados com alguns experimentos que utilizam os mesmos bancos de imagens. Os trabalho de (ALLAN et al., 2017) e (DUTTA et al., 2014) obtiveram 90% de acurácia utilizando o banco de imagens HRF, o trabalho atual utilizando a RNC DenseNet obteve 100% de acurácia. (ALLAN et al., 2017) obteve 86,4% utilizando o banco RIM-ONE-R3 enquanto a DenseNet obteve 95,4%. A acurácia média da DenseNet também foi superior aos demais trabalhos relacionados.

A Seção 5.1 apresenta algumas comparações adicionais entre as RNC. A rede SqueezeNet recebe destaque por possuir uma ótima acurácia e apresentar vantagens como:

- Tempo de processamento por época 13 vezes mais rápido.
- Consumo de memória quase 5 vezes menor.
- Menos da metade do tempo para realizar uma predição em uma imagem.
- Tamanho do modelo salvo pelo menos 15 vezes menor.

Essas vantagens possibilitam com que ela possa ser utilizada em sistemas embarcados e aplicativos para celular. O que auxilia a adoção em pequenos hospitais e clínicas, uma vez que não é tão necessário possuir um sistema dedicado para realizar a detecção em imagens.

5.2.1 Limitações do método proposto

Devido a baixa quantidade de bancos de imagens públicas devidamente anotados por especialistas, não é possível garantir uma boa generalização das RNC. Bancos de imagens como o HRF mesmo com técnicas de aumento de dados causam um grande sobreajuste nas RNC, prejudicando seu desempenho para a classificação de imagens no mundo real.

Outro grande problema é a falta de padronização das imagens do fundo do olho. Máquinas diferentes geram imagens diferentes. Isso combinado com formas de manuseio,

falta de manutenção e problemas de iluminação podem gerar imagens muito distintas das imagens na qual a RNC foi treinada, inutilizando o classificador, caso ele não tenha tido um treino robusto.

O capítulo a seguir apresenta as considerações finais e trabalhos futuros.

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho teve como objetivo realizar um comparativo entre a acurácia e outras propriedades de diversos modelos de RNC, com o propósito de auxiliar a implementação de futuros sistemas para auxiliar os profissionais responsáveis por realizar a detecção do glaucoma manualmente. A detecção do glaucoma em imagens do fundo do olho é realizada através da análise da RI, que pode conter certas características que acusam a presença do glaucoma. Existem trabalhos relacionados com foco na detecção da RI e com foco na detecção do glaucoma, assim como existem trabalhos que realizam as duas etapas. Muitos dos trabalhos relacionados não utilizavam bancos de imagens totalmente públicos, então os experimentos foram realizados utilizando bancos de imagens públicos como: HRF, RIM-ONE-R1, RIM-ONE-R2 e RIM-ONE-R3. Após algumas análises nos bancos, os bancos RIM-ONE-R1 e RIM-ONE-R2 foram fundidos formando o RIM-ONE-R1-R2. Utilizar esses bancos de imagem possibilita com que trabalhos futuros possam comparar seus resultados com os gerados nesse trabalho. Muitos dos trabalhos relacionados para a detecção da RI utilizam a intensidade dos pixeis para encontrar essa região, esse método pode não conseguir detectar a RI com precisão caso a imagem possua muitos artefatos luminosos ou ruido. O método de janela deslizante utilizado por (ALLAN et al., 2017) não foi utilizado por possuir problemas de performance, ao aplicar o classificador muitas vezes por imagem. O método de (CHAI et al., 2017) utiliza uma rede especializada em detectar objetos, obtendo uma performance melhor do que a do método anterior ao realizar a detecção. Para esse trabalho também foi utilizado uma rede especializada em detectar objetos, chamada de YOLO9000, que possui grandes ganhos de performance, segundo os experimentos do autor, em relação ao método anterior. Para a detecção do glaucoma os trabalhos relacionados em sua grande maioria utilizam RNC. Esse trabalho faz o comparativo de vários outros modelos de RNC para verificar se existe alguma rede com maior acurácia. Foi concluído que a DenseNet apresenta maior acurácia média do que as utilizadas pelos trabalhos relacionados. Os trabalho de (ALLAN et al., 2017) e (DUTTA et al., 2014) obtiveram 90% de acurácia utilizando o banco de imagens HRF, o trabalho atual utilizando a RNC DenseNet obteve 100% de acurácia. (ALLAN et al., 2017) obteve 86,4% utilizando o banco RIM-ONE-R3 enquanto a DenseNet obteve 95,4%.

Trabalhos futuros podem ser realizados utilizando diferentes bancos de imagens e modelos de RNC. Também é possível aplicar a rede SqueezeNet em projetos para dispositivos móveis envolvendo a detecção de glaucoma. Também objetiva-se desenvolver um modelo de arquitetura que apresente uma quantidade menor de parâmetros para compensar a quantidade de imagens públicas disponíveis.

REFERÊNCIAS BIBLIOGRÁFICAS

ABADI, M. et al. TensorFlow: A system for large-scale machine learning. *USENIX Symposium on Operating Systems Design and Implementation*, abs/1605.0, 2016. Disponível em: <http://arxiv.org/abs/1605.08695>.

ALLAN, C. et al. Automatic identification of glaucoma using deep learning methods. *Studies in Health Technology and Informatics*, IOS Press, v. 245, n. MEDINFO 2017: Precision Healthcare through Informatics, p. 318–321, 2017. ISSN 0926-9630. Disponível em: <http://doi.org/10.3233/978-1-61499-830-3-318>.

Anderson Vinicius. *Redes Neurais Artificiais Fundamentos teóricos*. Medium, 2017. Acessado em 11 dez 2017. Disponível em: <https://medium.com/@avinicius.adorno/redes-neurais-artificiais-5b65a43614a0>.

Andrej Karpathy. *Convolutional Neural Networks for Visual Recognition*. 2017. Acessado em 01 jan 2018. Disponível em: <http://cs231n.github.io/convolutional-networks/>.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, n. 8, p. 1798–1828, 2013. ISSN 01628828.

BLOICE, M. D.; STOCKER, C.; HOLZINGER, A. Augmentor: An image augmentation library for machine learning. *CoRR*, abs/1708.04680, 2017. Disponível em: <http://arxiv.org/abs/1708.04680>.

CERDA, J.; CIFUENTES, L. Uso de curvas ROC en investigación clínica. Aspectos teórico-prácticos. *Revista chilena de infectología*, scielocl, v. 29, p. 138 – 141, 04 2012. ISSN 0716-1018. Disponível em: <https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0716-10182012000200003&nrm=iso>.

CHAI, Y. et al. Deep learning through two-branch convolutional neuron network for glaucoma diagnosis. In: ____. *Smart Health: International Conference, ICSH 2017, Hong Kong, China, June 26-27, 2017, Proceedings*. Cham: Springer International Publishing, 2017. p. 191–201. ISBN 978-3-319-67964-8. Disponível em: <https://doi.org/10.1007/978-3-319-67964-8_19>.

CHEN, X. et al. Glaucoma detection based on deep convolutional neural network. p. 715–718, Aug 2015. ISSN 1094-687X.

DUTTA, M. K. et al. Glaucoma detection by segmenting the super pixels from fundus colour retinal images. In: *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*. [S.l.: s.n.], 2014. p. 86–90.

Eduardo Prado. *Reconhecimento de imagens: Um novo aliado do diagnóstico Digital na Medicina*. 2016. Acessado em 18 dez 2017. Disponível em: <https://www.ibm.com/blogs/robertoa/2016/03/reconhecimento-de-imagens-um-novo-aliado-do-diagnostico-digital-na-medicina/>.

FUMERO, F. et al. RIM-ONE: An open retinal image database for optic nerve evaluation. In: *Proceedings - IEEE Symposium on Computer-Based Medical Systems*. [S.l.: s.n.], 2011. ISBN 9781457711909. ISSN 10637125.

Girshick, R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. *ArXiv e-prints*, nov. 2013.

Glaucoma Research Foundation. *Five Common Glaucoma Tests*. 2017. Acessado em 29 dez 2017. Disponível em: <https://www.glaucoma.org/glaucoma/diagnostic-tests.php>.

HE, K. et al. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Disponível em: <http://arxiv.org/abs/1512.03385>.

Huang, G. et al. Densely Connected Convolutional Networks. *ArXiv e-prints*, ago. 2016.

IANDOLA, F. N. et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size. *arXiv:1602.07360*, 2016.

KOLAR, R. et al. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Processing*, v. 7, n. 4, p. 373–383, 2013. ISSN 1751-9659. Disponível em: <http://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2012.0455>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012. p. 1097–1105. Disponível em: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.

LALONDE, M.; BEAULIEU, M.; GAGNON, L. Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching. *IEEE Transactions on Medical Imaging*, v. 20, n. 11, p. 1193–1200, Nov 2001. ISSN 0278-0062.

MATSUGU, M. et al. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, v. 16, n. 5, p. 555 – 559, 2003. ISSN 0893-6080. Advances in Neural Networks Research: IJCNN '03. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0893608003001151>.

National Eye Institute. *Facts About Glaucoma*. 2014. Acessado em 06 dez 2017. Disponível em: <https://nei.nih.gov/health/glaucoma/glaucoma_facts>.

ORLANDO, J. I. et al. Convolutional neural network transfer for automated glaucoma identification. *Proc.SPIE*, v. 10160, p. 10160 – 10160 – 10, 2017. Disponível em: <http://dx.doi.org/10.1117/12.2255740>.

PASZKE, A. et al. Automatic differentiation in pytorch. 2017.

QUIGLEY, H. A. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, BMJ, v. 90, n. 3, p. 262–267, mar 2006. Disponível em: <https://doi.org/10.1136/bjo.2005.081224>.

RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, v. 29, n. 9, p. 2352–2449, 2017. PMID: 28599112. Disponível em: <https://doi.org/10.1162/neco_a_00990>.

Redmon, J. et al. You Only Look Once: Unified, Real-Time Object Detection. *ArXiv e-prints*, jun. 2015.

REDMON, J.; FARHADI, A. YOLO9000: Better, Faster, Stronger. *arXiv preprint arXiv:1612.08242*, 2016. Disponível em: <http://arxiv.org/abs/1612.08242>.

SCHMIDHUBER, J. Deep learning in neural networks : An overview. *Neural Networks journal*, v. 61, p. 85–117, 2015. Disponível em: <http://arxiv.org/abs/1404.7828>.

SERMANET, P. et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. *arXiv preprint arXiv*, p. 1312.6229, 2013. Disponível em: <http://arxiv.org/abs/1312.6229>.

SHEEBA, O. et al. Glaucoma Detection Using Artificial Neural Network. *IACSIT International Journal of Engineering and Technology*, v. 6, n. 2, p. 158–161, 2014. ISSN 17938236.

Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv e-prints*, set. 2014.

SINTHANAYOTHIN, C. et al. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophthalmology*, BMJ Publishing Group Ltd, v. 83, n. 8, p. 902–910, 1999. ISSN 0007-1161. Disponível em: <http://bjo.bmj.com/content/83/8/902>.

Szegedy, C. et al. Rethinking the Inception Architecture for Computer Vision. *ArXiv e-prints*, dez. 2015.

THAM, Y.-C. et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040. *Ophthalmology*, Elsevier BV, v. 121, n. 11, p. 2081–2090, nov 2014. Disponível em: <https://doi.org/10.1016/j.ophtha.2014.05.013>.

Ujjwal Karn. *An Intuitive Explanation of Convolutional Neural Networks*. 2016. Acessado em 01 jan 2018. Disponível em: <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>.

WALTER, T.; KLEIN, J.-C. Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques. In: ____. *Medical Data Analysis: Second International Symposium, ISMDA 2001 Madrid, Spain, October 8–9, 2001 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 282–287. ISBN 978-3-540-45497-7. Disponível em: <https://doi.org/10.1007/3-540-45497-7_43>.

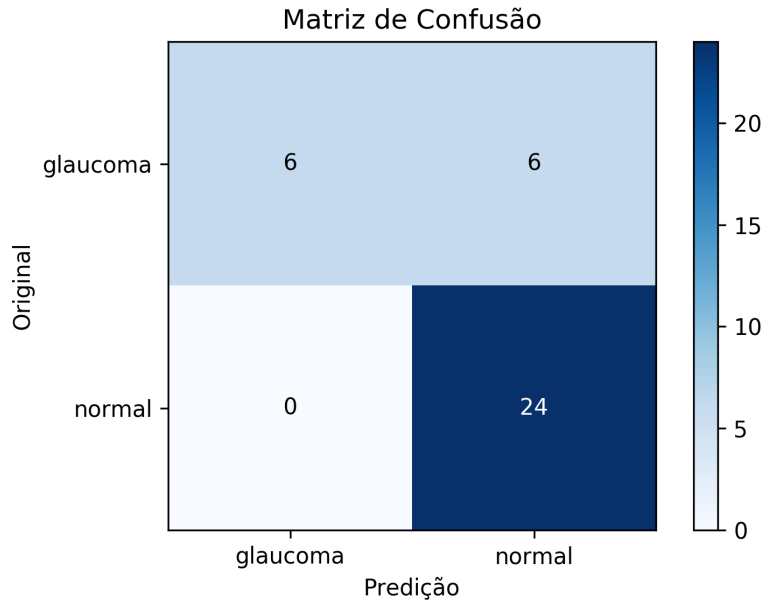
WILSON, M. Intraocular pressure: Does it measure up? *The Open Ophthalmology Journal*, Bentham Science Publishers Ltd., v. 3, n. 1, p. 32–37, aug 2009. Disponível em: <https://doi.org/10.2174/1874364100903010032>.

XU, Y. et al. Sliding Window and Regression Based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, v. 14 Pt 3, p. 1–8, 2011.

APÊNDICE A – RESULTADOS DE CADA EXPERIMENTO

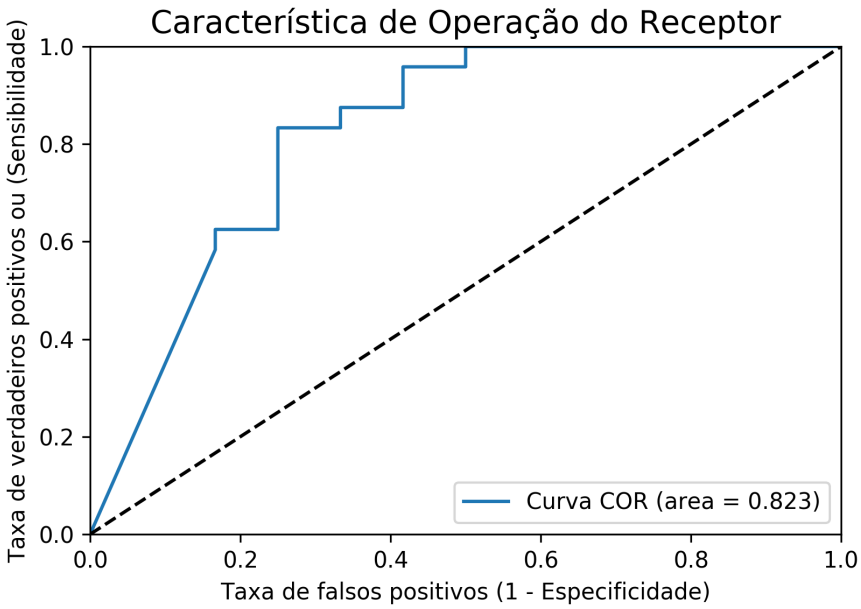
A.1 ALEXNET

Figura A.1 – Matriz de Confusão obtida na validação da rede AlexNet utilizando o banco de imagens HRF.



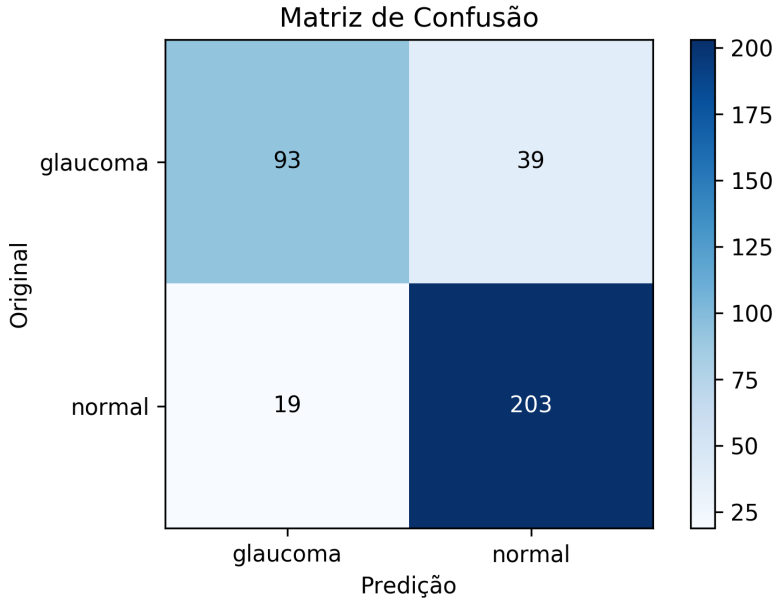
Fonte: Autoria própria.

Figura A.2 – Curva COR obtida na validação da rede AlexNet utilizando o banco de imagens HRF.



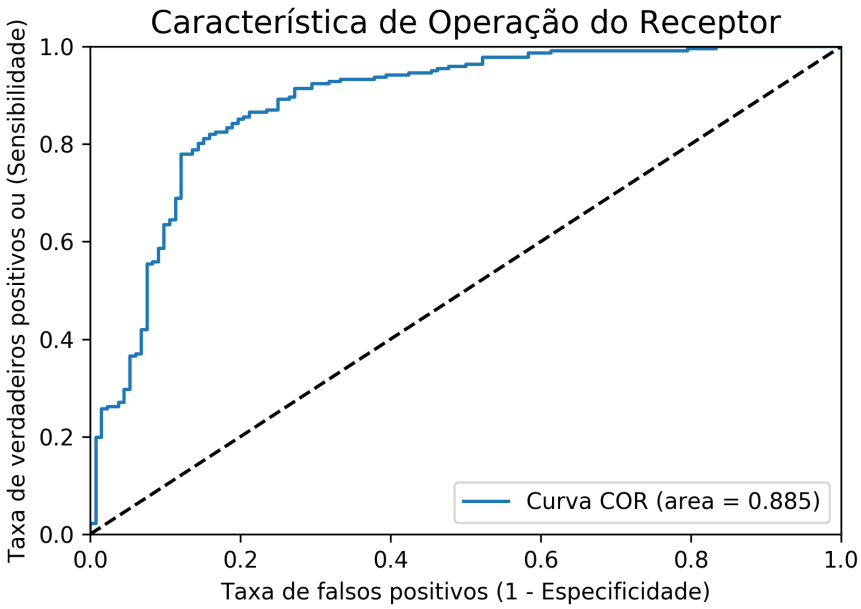
Fonte: Autoria própria.

Figura A.3 – Matriz de Confusão obtida na validação da rede AlexNet utilizando o banco de imagens R1-R2.



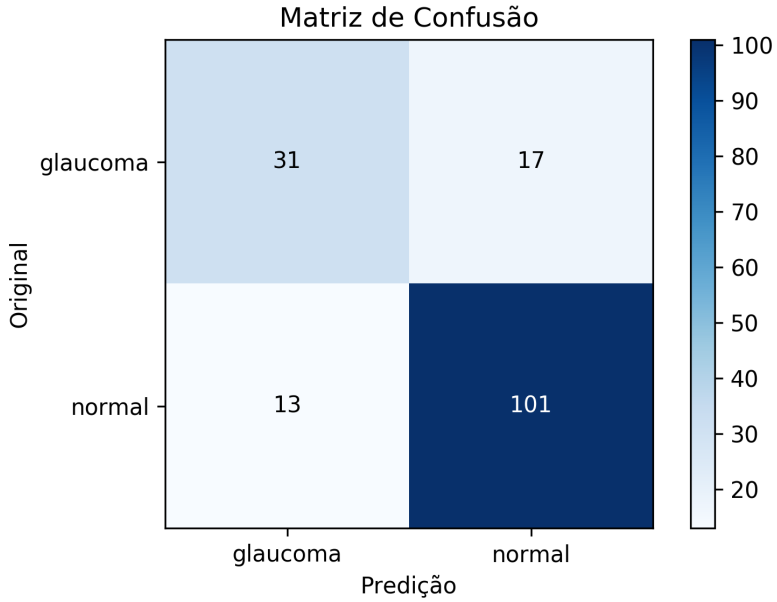
Fonte: Autoria própria.

Figura A.4 – Curva COR obtida na validação da rede AlexNet utilizando o banco de imagens R1-R2.



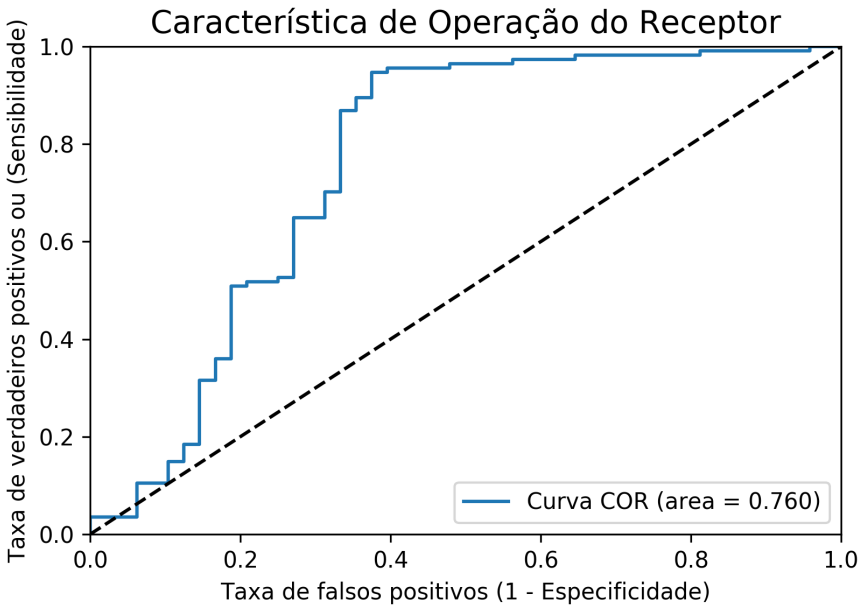
Fonte: Autoria própria.

Figura A.5 – Matriz de Confusão obtida na validação da rede AlexNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

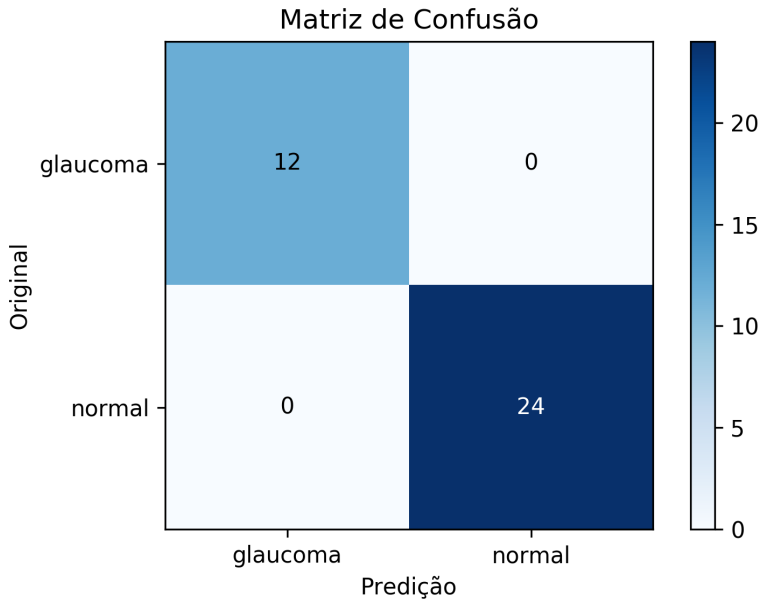
Figura A.6 – Curva COR obtida na validação da rede AlexNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

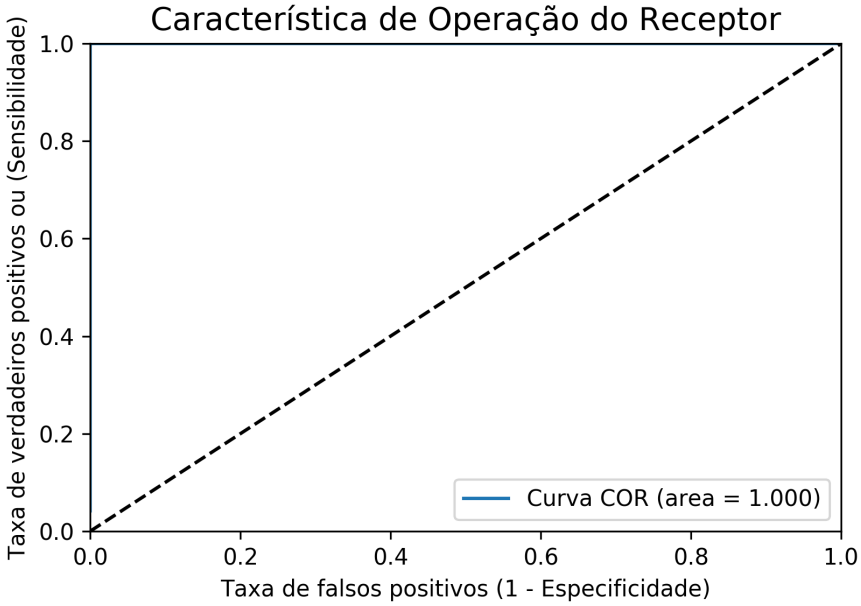
A.2 DENSENET

Figura A.7 – Matriz de Confusão obtida na validação da rede DenseNet utilizando o banco de imagens HRF.



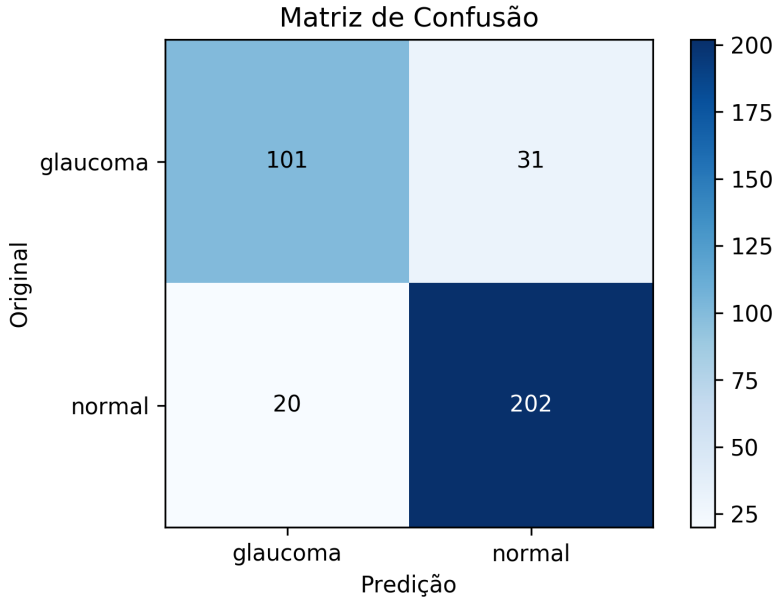
Fonte: Autoria própria.

Figura A.8 – Curva COR obtida na validação da rede DenseNet utilizando o banco de imagens HRF.



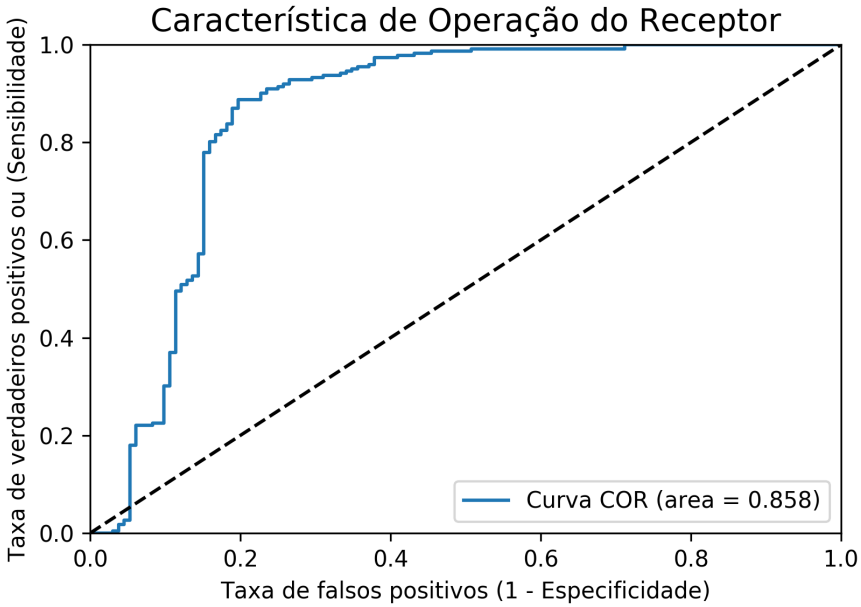
Fonte: Autoria própria.

Figura A.9 – Matriz de Confusão obtida na validação da rede DenseNet utilizando o banco de imagens R1-R2.



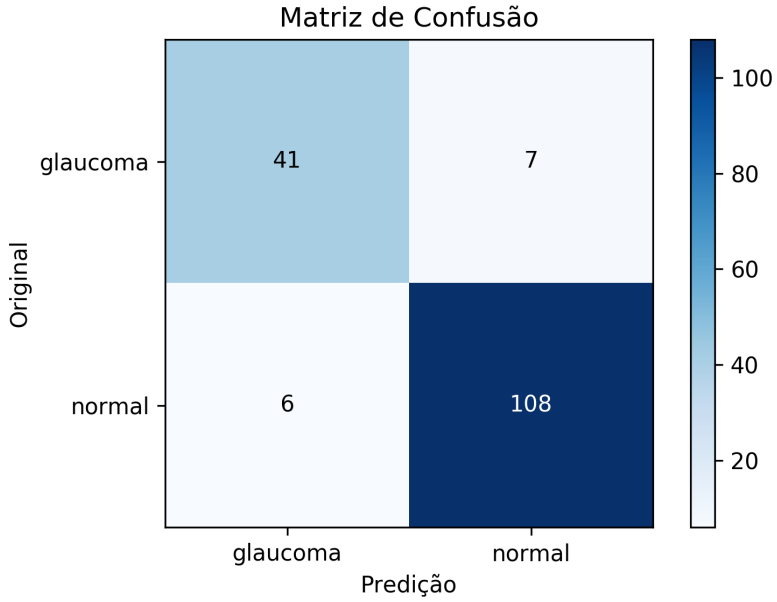
Fonte: Autoria própria.

Figura A.10 – Curva COR obtida na validação da rede DenseNet utilizando o banco de imagens R1-R2.



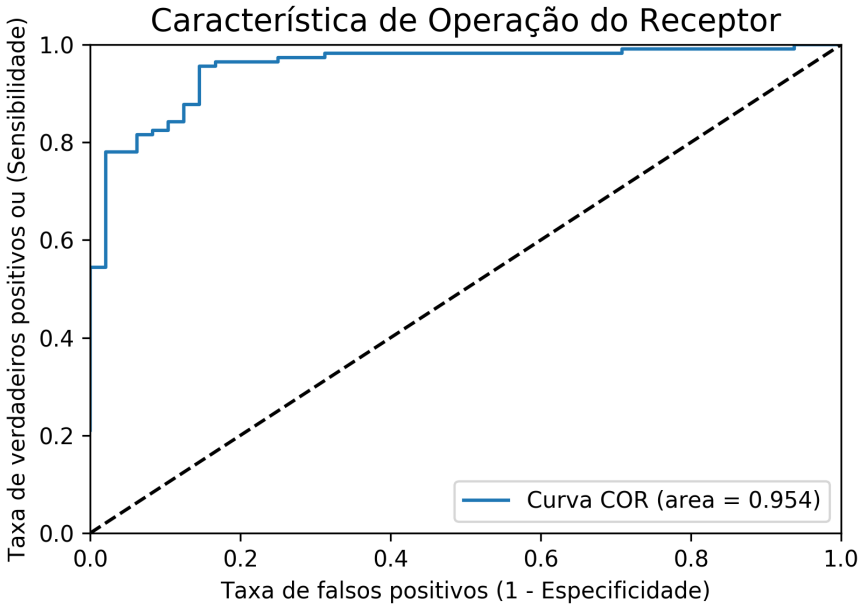
Fonte: Autoria própria.

Figura A.11 – Matriz de Confusão obtida na validação da rede DenseNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

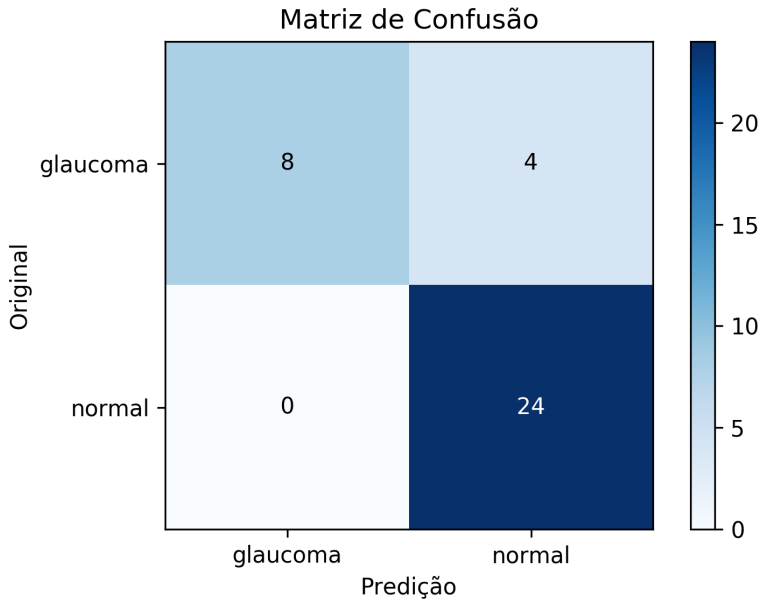
Figura A.12 – Curva COR obtida na validação da rede DenseNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

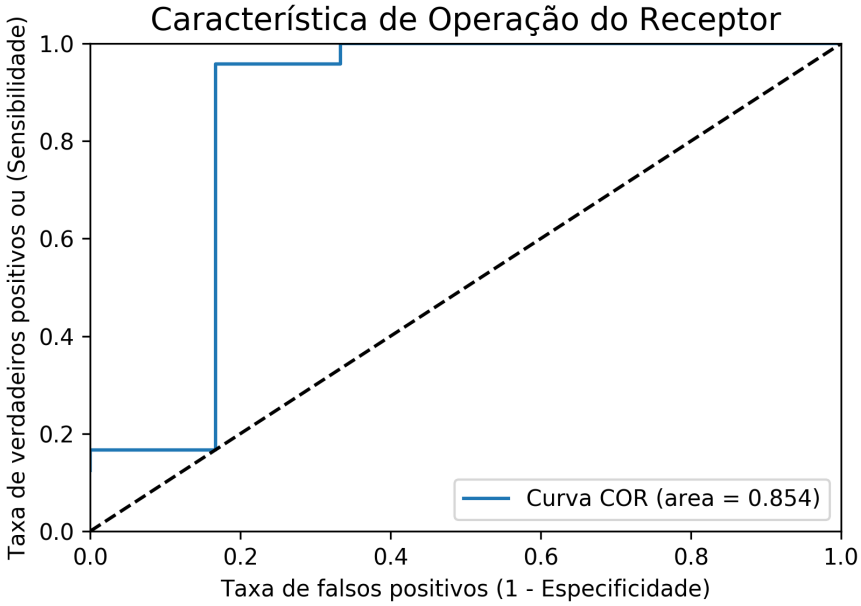
A.3 INCEPTION

Figura A.13 – Matriz de Confusão obtida na validação da rede Inception utilizando o banco de imagens HRF.



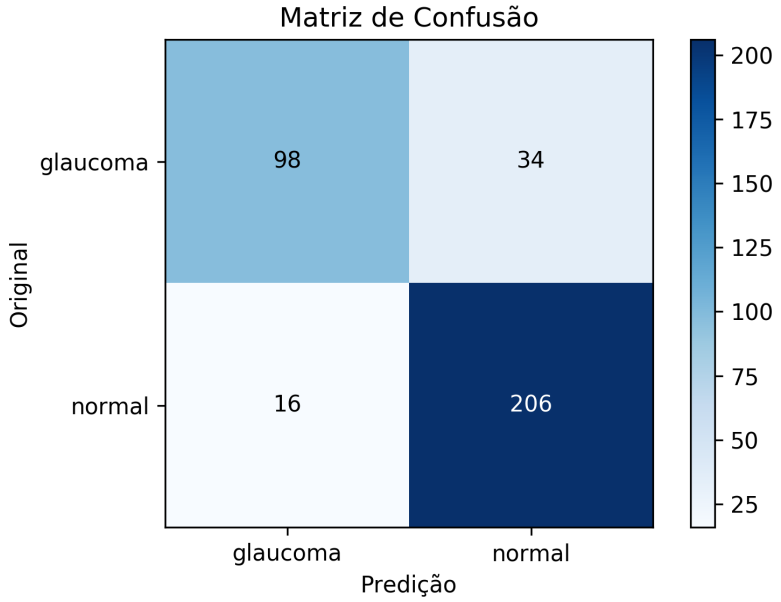
Fonte: Autoria própria.

Figura A.14 – Curva COR obtida na validação da rede Inception utilizando o banco de imagens HRF.



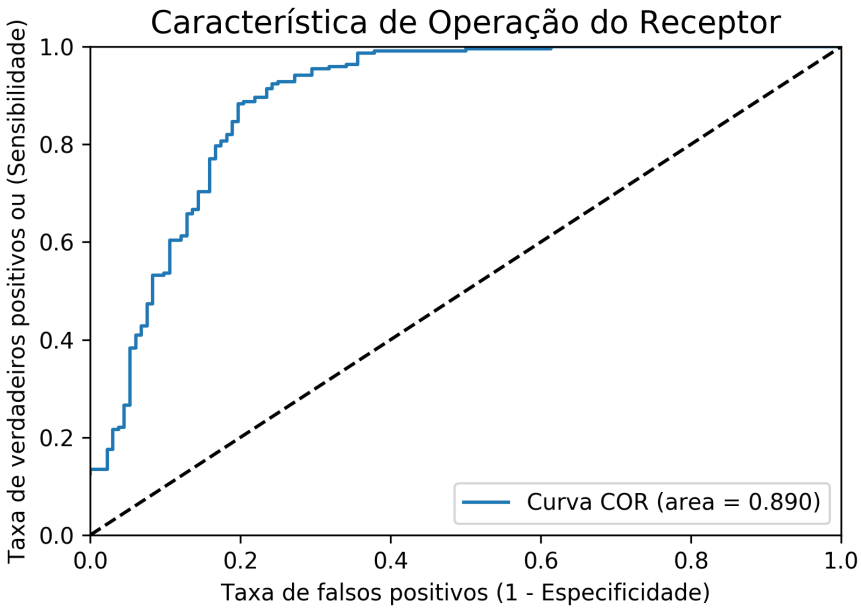
Fonte: Autoria própria.

Figura A.15 – Matriz de Confusão obtida na validação da rede Inception utilizando o banco de imagens R1-R2.



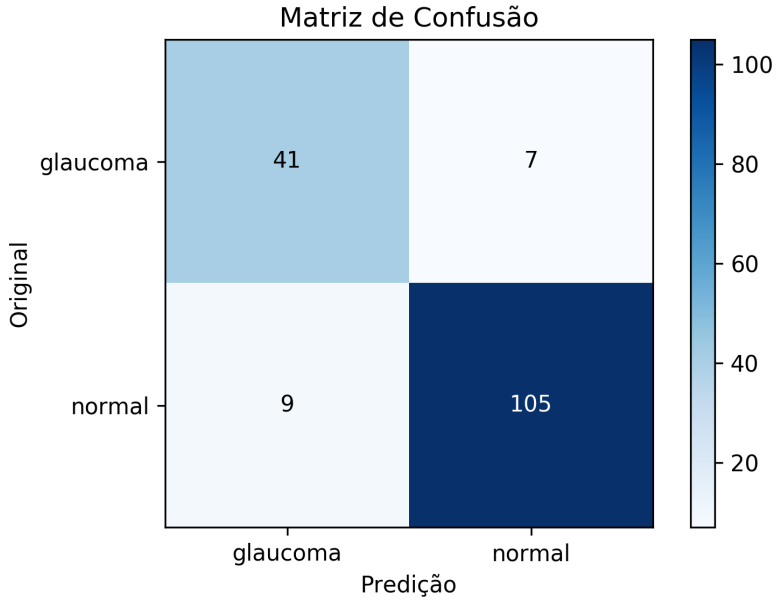
Fonte: Autoria própria.

Figura A.16 – Curva COR obtida na validação da rede Inception utilizando o banco de imagens R1-R2.



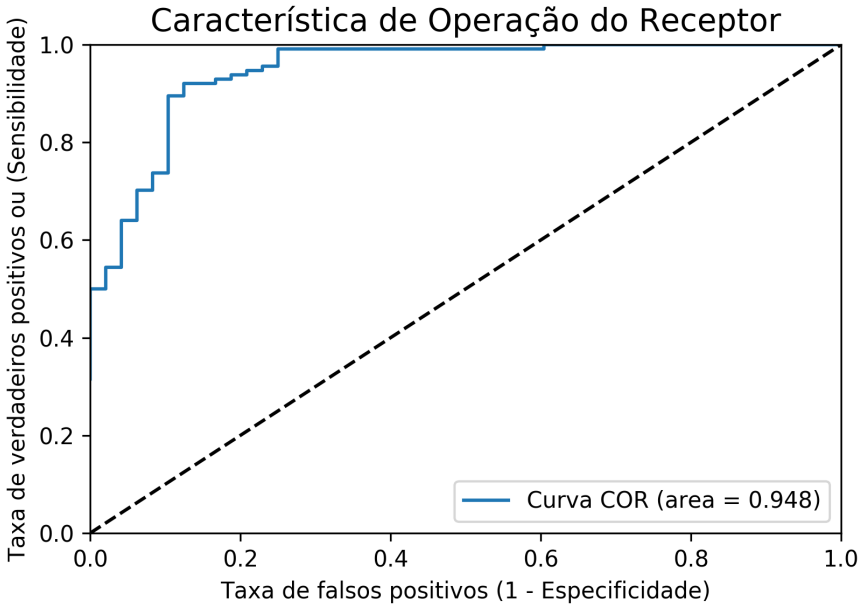
Fonte: Autoria própria.

Figura A.17 – Matriz de Confusão obtida na validação da rede Inception utilizando o banco de imagens R3.



Fonte: Autoria própria.

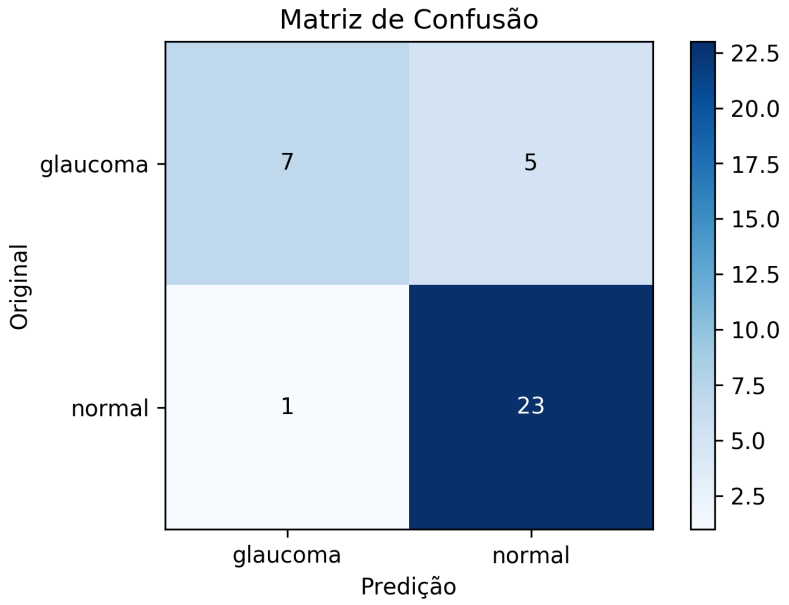
Figura A.18 – Curva COR obtida na validação da rede Inception utilizando o banco de imagens R3.



Fonte: Autoria própria.

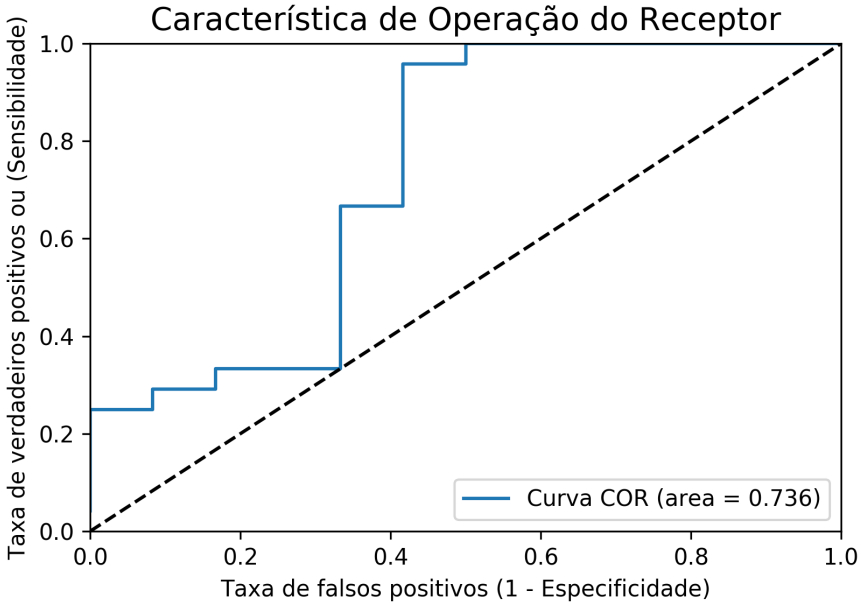
A.4 RESNET

Figura A.19 – Matriz de Confusão obtida na validação da rede ResNet utilizando o banco de imagens HRF.



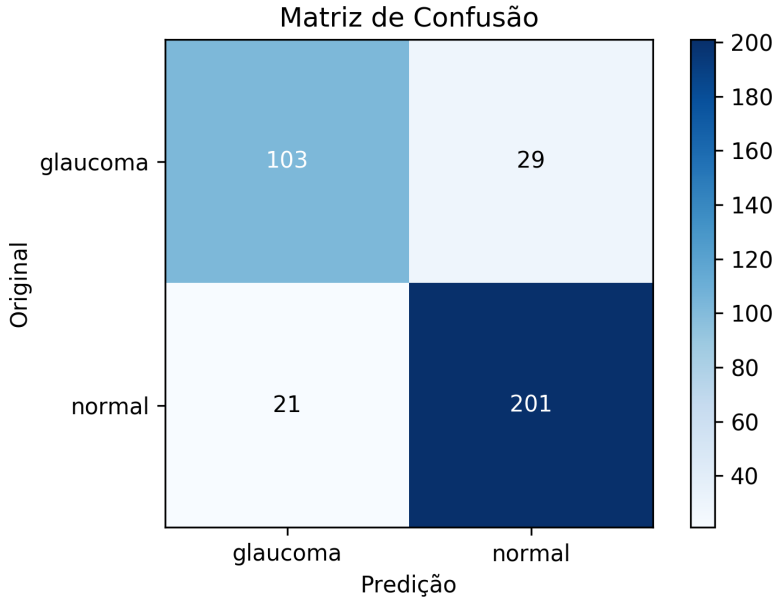
Fonte: Autoria própria.

Figura A.20 – Curva COR obtida na validação da rede ResNet utilizando o banco de imagens HRF.



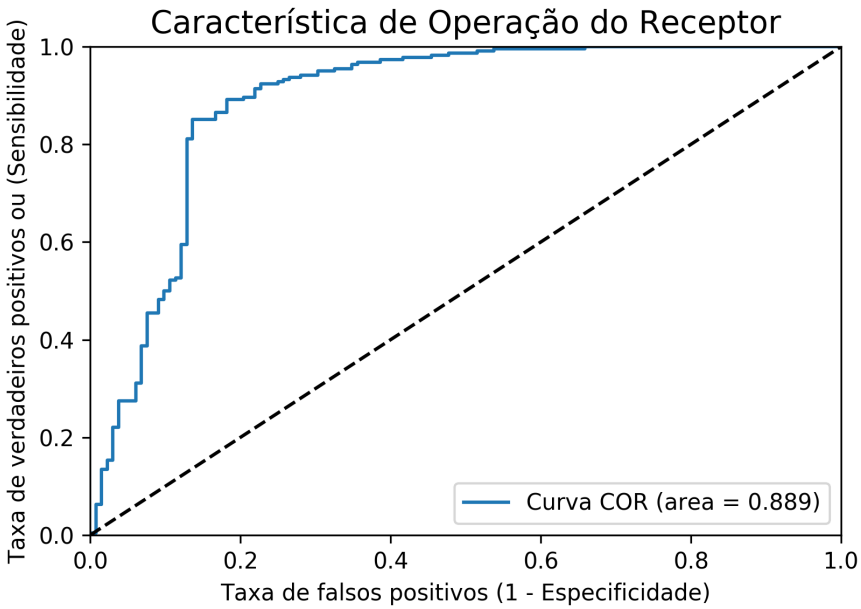
Fonte: Autoria própria.

Figura A.21 – Matriz de Confusão obtida na validação da rede ResNet utilizando o banco de imagens R1-R2.



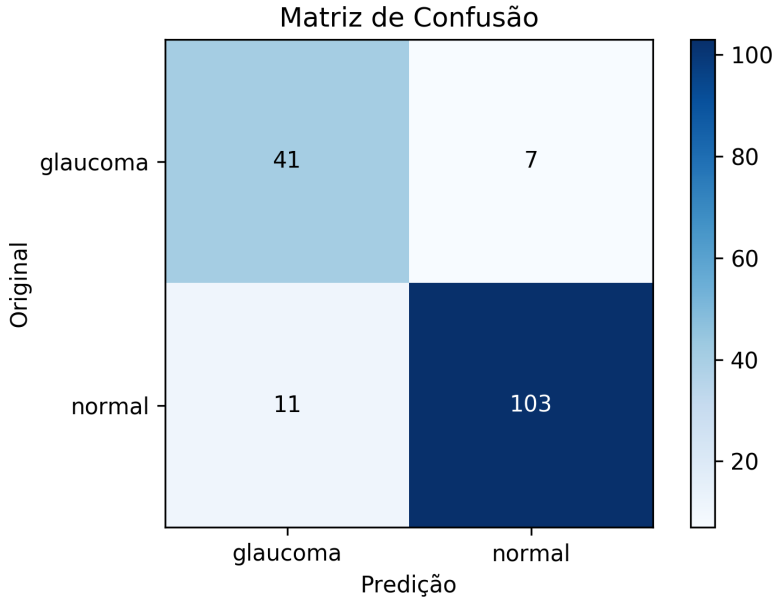
Fonte: Autoria própria.

Figura A.22 – Curva COR obtida na validação da rede ResNet utilizando o banco de imagens R1-R2.



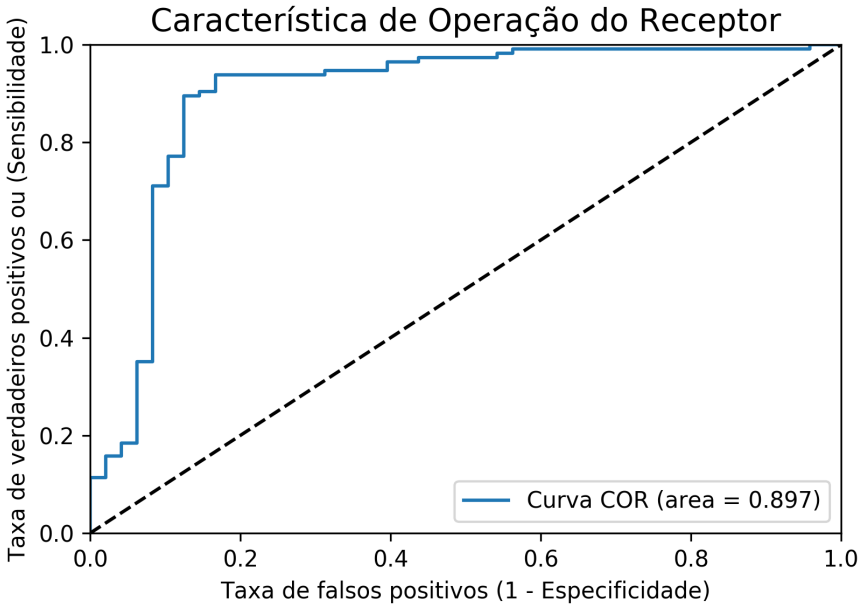
Fonte: Autoria própria.

Figura A.23 – Matriz de Confusão obtida na validação da rede ResNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

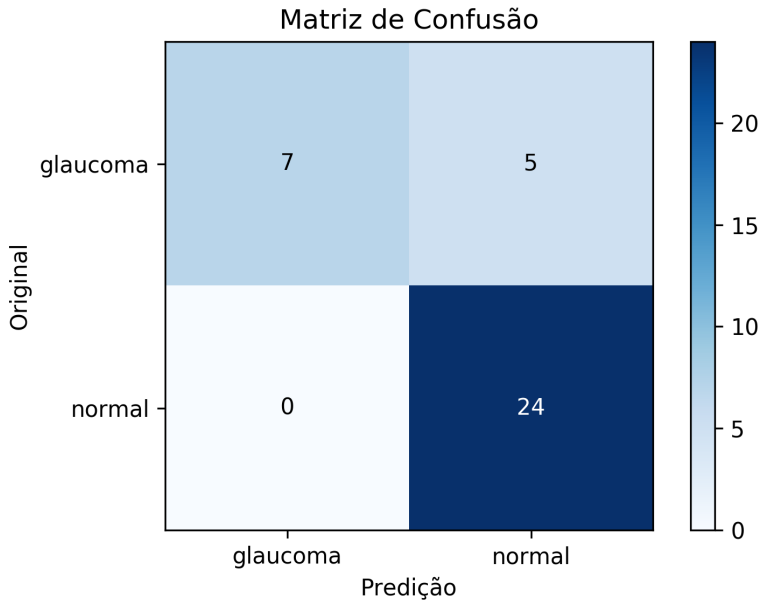
Figura A.24 – Curva COR obtida na validação da rede ResNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

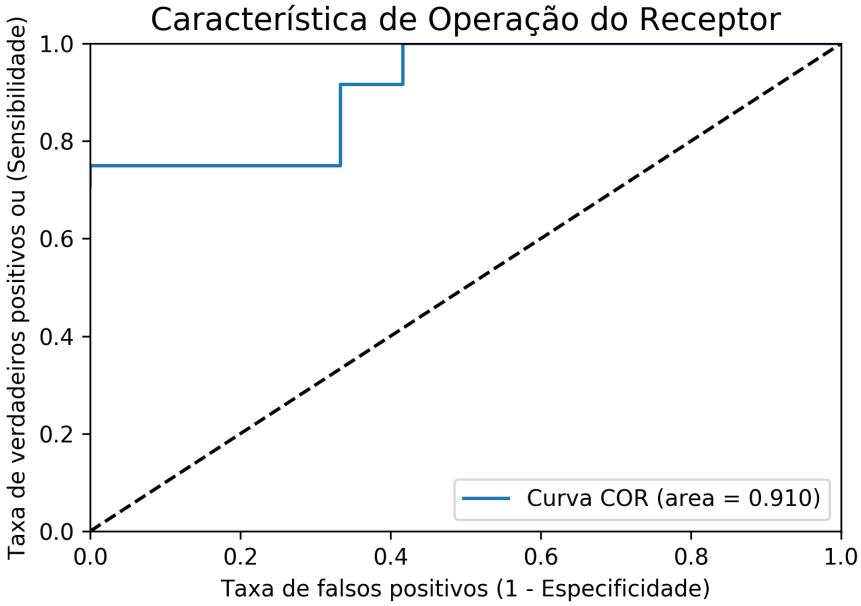
A.5 SQUEEZENET

Figura A.25 – Matriz de Confusão obtida na validação da rede SqueezeNet utilizando o banco de imagens HRF.



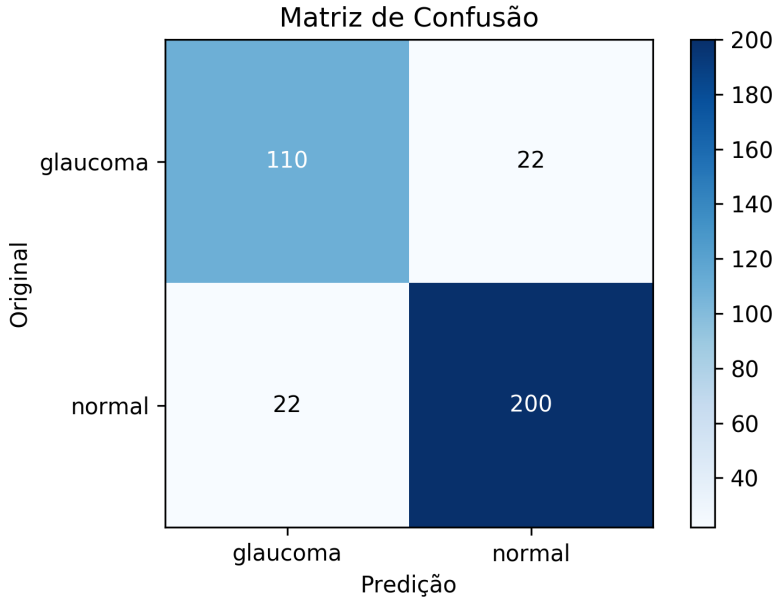
Fonte: Autoria própria.

Figura A.26 – Curva COR obtida na validação da rede SqueezeNet utilizando o banco de imagens HRF.



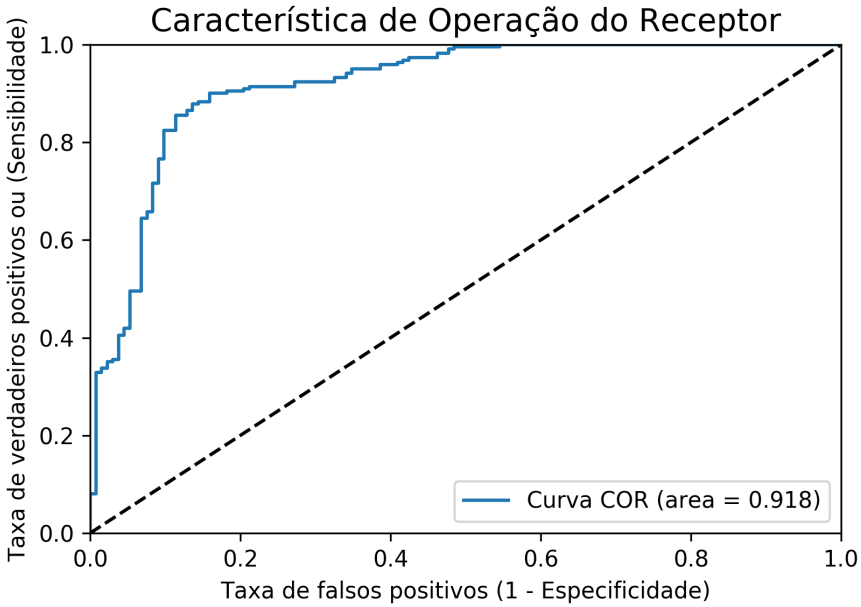
Fonte: Autoria própria.

Figura A.27 – Matriz de Confusão obtida na validação da rede SqueezeNet utilizando o banco de imagens R1-R2.



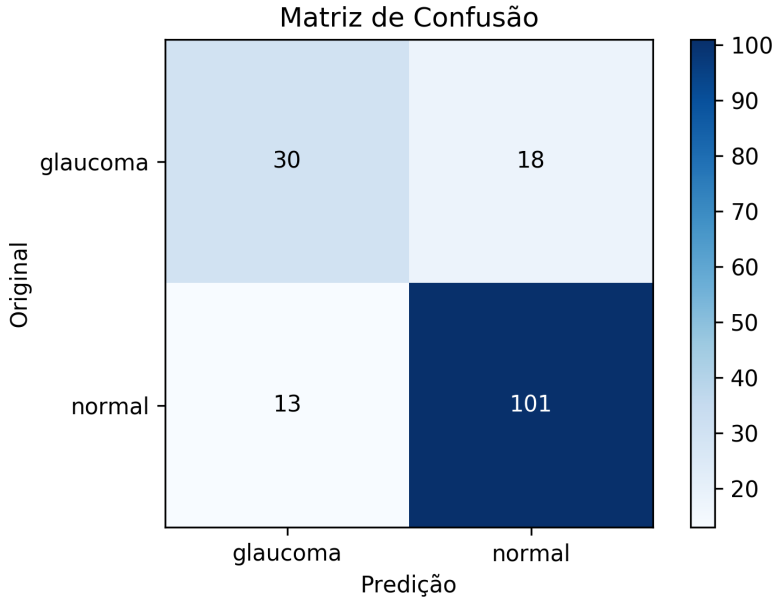
Fonte: Autoria própria.

Figura A.28 – Curva COR obtida na validação da rede SqueezeNet utilizando o banco de imagens R1-R2.



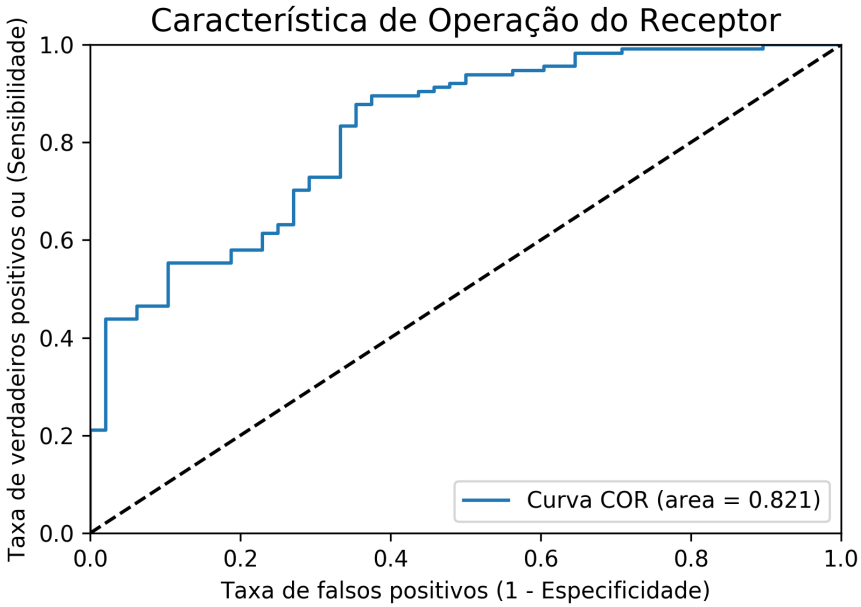
Fonte: Autoria própria.

Figura A.29 – Matriz de Confusão obtida na validação da rede SqueezeNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

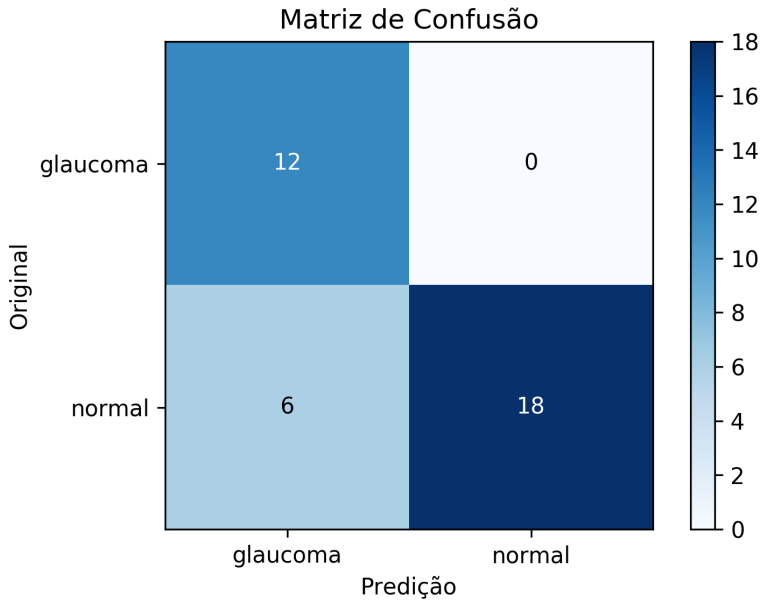
Figura A.30 – Curva COR obtida na validação da rede SqueezeNet utilizando o banco de imagens R3.



Fonte: Autoria própria.

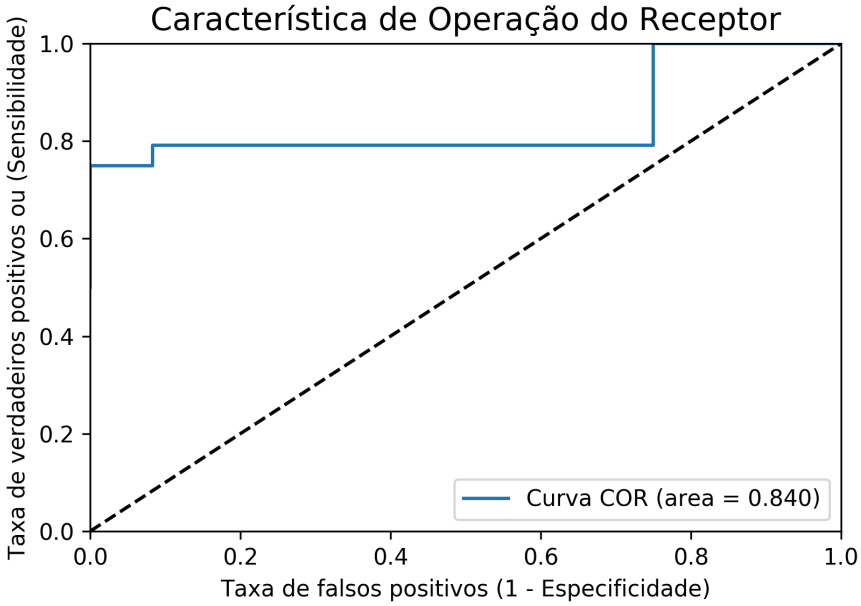
A.6 VGG-16

Figura A.31 – Matriz de Confusão obtida na validação da rede VGG-16 utilizando o banco de imagens HRF.



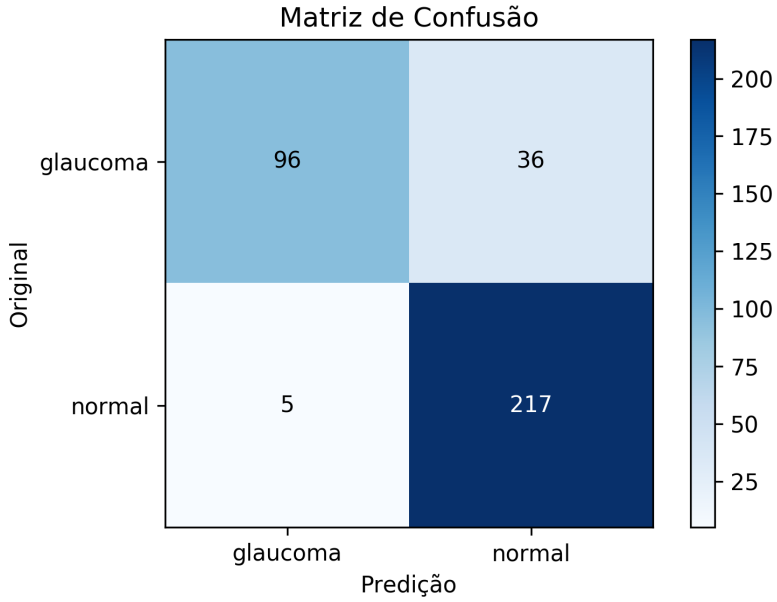
Fonte: Autoria própria.

Figura A.32 – Curva COR obtida na validação da rede VGG-16 utilizando o banco de imagens HRF.



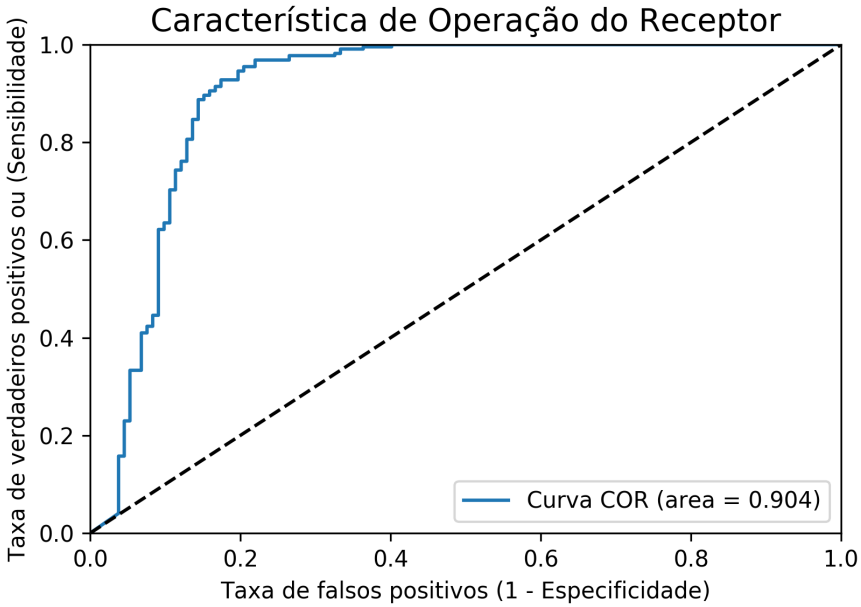
Fonte: Autoria própria.

Figura A.33 – Matriz de Confusão obtida na validação da rede VGG-16 utilizando o banco de imagens R1-R2.



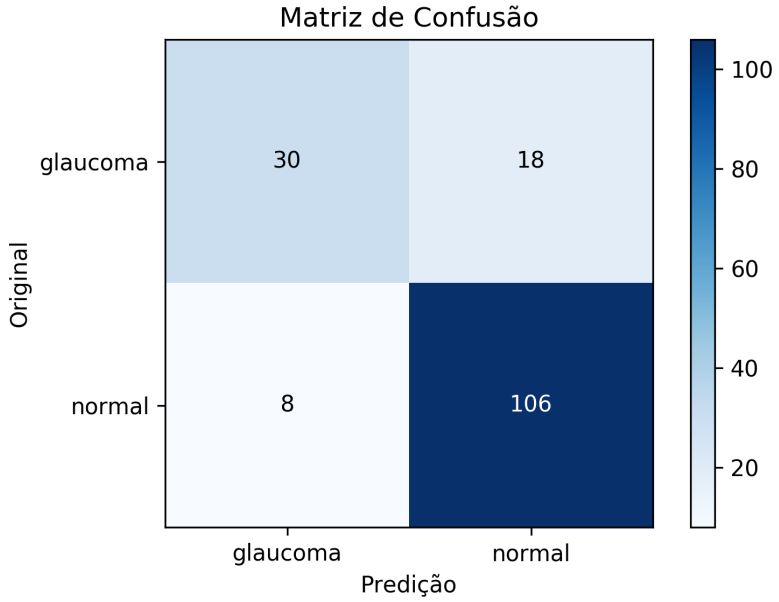
Fonte: Autoria própria.

Figura A.34 – Curva COR obtida na validação da rede VGG-16 utilizando o banco de imagens R1-R2.



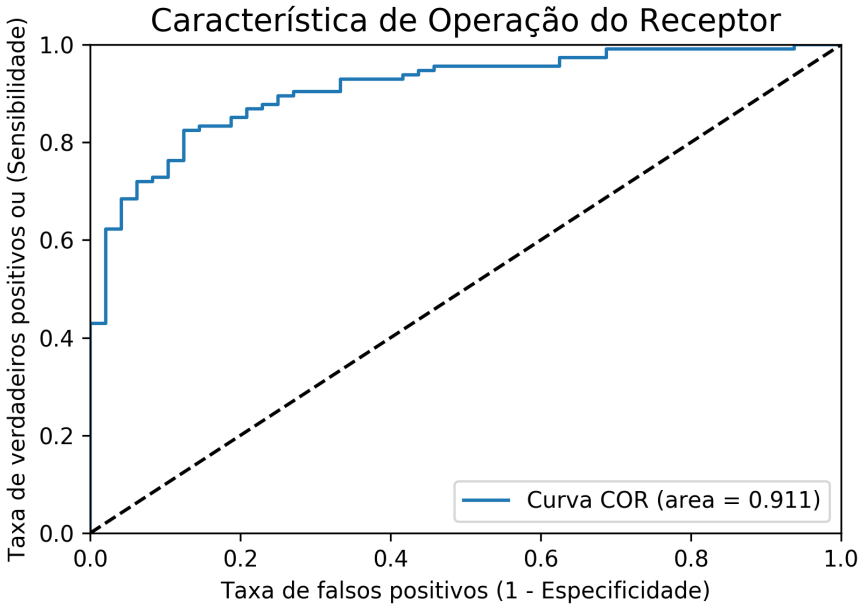
Fonte: Autoria própria.

Figura A.35 – Matriz de Confusão obtida na validação da rede VGG-16 utilizando o banco de imagens R3.



Fonte: Autoria própria.

Figura A.36 – Curva COR obtida na validação da rede VGG-16 utilizando o banco de imagens R3.



Fonte: Autoria própria.