

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS  
ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO  
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**SEGMENTAÇÃO DE VEIAS DO PULSO COM USO DE REDES NEURAIS  
CONVOLUCIONAIS**

MATHEUS SOUZA CARDOSO

GOIÂNIA  
2020

MATHEUS SOUZA CARDOSO

**SEGMENTAÇÃO DE VEIAS DO PULSO COM USO DE REDES NEURAIAS  
CONVOLUCIONAIS**

Trabalho de Conclusão de Curso apresentado à Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Ciência da Computação.

Orientador:

Prof. Dr. Arlindo Rodrigues Galvão Filho

Banca examinadora:

Prof. Dr. Rafael Viana de Carvalho

Prof. Dr. Ricardo Augusto Pereira Franco

GOIÂNIA  
2020

MATHEUS SOUZA CARDOSO

**SEGMENTAÇÃO DE VEIAS DO PULSO COM USO DE REDES NEURAIS  
CONVOLUCIONAIS**

Este Trabalho de Conclusão de Curso julgado adequado para obtenção do título de Bacharel em Ciência da Computação, e aprovado em sua forma final pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, em \_\_\_\_/\_\_\_\_/\_\_\_\_\_.

---

Prof. Me. Ludmilla Reis Pinheiro dos Santos  
Coordenadora de Trabalho de Conclusão de  
Curso

Banca examinadora:

---

Orientador: Prof. Dr. Arlindo Rodrigues Galvão  
Filho

---

Prof. Dr. Rafael Viana de Carvalho

---

Prof. Dr. Ricardo Augusto Pereira Franco

GOIÂNIA  
2020

## RESUMO

Este trabalho propõe o estudo de segmentação semântica de imagens utilizando redes neurais profundas. Foi gerada uma base de dados de anotações de imagens do pulso no espectro infravermelho, para realizar o treinamento de diferentes Redes Neurais Convolucionais (CNNs) para segmentação semântica. Em um primeiro momento foram treinados três CNNs diferentes (U-Net, SegNet e DeepLabv3) utilizando a mesma disposição de dados de treinamento e teste, onde as imagens dos primeiros 56 indivíduos foram utilizadas para treinamento e as últimas 37 para teste, os resultados dos modelos foram comparados e a que obteve os melhores valores de acurácia para detecção de veias foi a SegNet. Por fim o modelo SegNet foi treinado novamente com a separação dos dados de treinamento e teste feito de duas formas diferentes, no primeiro caso foram utilizadas três imagens de cada indivíduo para treinamentos e duas para teste e no segundo caso os dados de treinamento e teste foram distribuídos aleatoriamente, sendo 60% das imagens totais para treinamento e 40% para teste, para estes dois últimos testes não houve melhora nos resultados. Foi possível concluir que a SegNet se mostrou ser mais eficaz para a segmentação de veias para a base de dados utilizada no trabalho em comparação com outras CNNs, além disso, foi possível observar que os modelos em geral não obtiveram excelentes resultados, pois foram utilizados poucos dados para treinamento e teste.

**Palavras-Chave:** Redes Neurais Convolucionais, Veias, Segmentação Semântica, Base de Dados.

## ABSTRACT

This work proposes the study of image semantic verification using deep neural networks. A database of wrist image annotations in the infrared spectrum was generated to train different Convolutional Neural Networks (CNNs) for supra-semantic. At first, three different CNNs (U-Net, SegNet and DeepLabv3) were trained using the same training and test data arrangement, where images of the first 56 individuals were used for training and the last 37 for testing, the the results of the models were compared and the one that obtained the best accuracy values for vein detection was SegNet. Finally, the SegNet model was re-trained with the separation of the training and test data done in two different ways, in the first case three images of each individual were used for training and two for testing and in the second case the training and test data were randomly distributed, with 60% of the total images for training and 40% for testing, for these last two tests there was no improvement in results. It was possible to conclude that a SegNet proved to be more effective for a combination of veins for the database to use non-work compared to other CNNs, in addition it was possible to observe that the models in general did not obtain excellent results, since only data were used for training and testing.

**Key words:** Convolutional Neural Networks, Veins, Sematic Segmentation, Database.

## SÚMARIO

<b>1. INTRODUÇÃO</b> .....	<b>7</b>
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>8</b>
2.1. Redes Neurais Convolucionais ( <i>Convolutional Neural Network</i> ).....	8
2.2. U-Net.....	10
2.3. DeepLabv3.....	10
2.4. SegNet .....	11
2.5. Anotação de imagem ( <i>Image Annotation</i> ).....	11
<b>3. MATERIAIS E MÉTODOS</b> .....	<b>12</b>
3.1. Estação de Aquisição.....	12
3.2. Base de Dados.....	13
3.3. Pré-processamento Digital .....	15
3.4. Construção do Modelo e Simulação.....	16
<b>4. RESULTADOS</b> .....	<b>18</b>
<b>5. CONCLUSÃO</b> .....	<b>19</b>
<b>REFERÊNCIAS</b> .....	<b>21</b>

## 1. INTRODUÇÃO

Com os avanços da tecnologia, imagens digitais vem sendo cada vez mais utilizadas em diversas áreas como em redes sociais, exames médicos, sites na *web*, entre outros. Tais imagens podem ser obtidas através de câmeras digitais instaladas em diversas estações de captura como, por exemplo, em radares de velocidade, satélites, celulares, etc.

Através de alguns equipamentos como, por exemplo de ressonância magnética, médicos podem realizar a aquisição de imagens digitais de boa qualidade de órgãos internos do corpo, incluindo tecidos ou vasos sanguíneos (WANG; LEEDHAM, 2006).

Além de adquirir imagens no espectro visível algumas câmeras como as multiespectrais conseguem realizar a aquisição de imagens em outros espectros (JENSEN, 1986) como o espectro infravermelho próximo (NIR). Com o espectro NIR é possível obter características adicionais anteriormente limitadas ao espectro do visível (KLARE; JAIN, 2010).

Um exemplo de aplicação deste tipo de imagem é a utilização de imagens NIR para auxiliar médicos e enfermeiras a encontrar com mais facilidade veias superficiais e então injetar drogas de forma mais precisa na veia do paciente (CHENG-TANG et al., 2019), isso é possível porque o NIR é absorvido pela hemoglobina que é uma proteína presente nos glóbulos vermelhos do sangue, com isso imagens com destaque em veias mais superficiais podem ser geradas.

Outra área que vem se destacando e desenvolvendo novas tecnologias com o uso de imagens NIR é a da biometria, que se baseia na geometria das veias da mão para criar modelos de reconhecimento de indivíduos (AGARWAL; MAHESHWARI; YADAV, 2014). Este método utiliza o padrão de distribuição dos vasos sanguíneos da mão em imagens no espectro NIR como fator de identificação. Tal modelo só realiza a autenticação se for identificado fluxo sanguíneo nas veias analisadas, fazendo com que possua uma confiabilidade elevada, pois dessa forma um indivíduo morto não pode ser identificado (SARKAR I. et al, 2010; KUMARI; ALANKAR; GROVER, 2016; BHOKARE et al. 2013).

No entanto, as imagens NIR geradas por câmeras multiespectrais, geralmente necessitam de algum pré-processamento para segmentação das veias, afim de serem utilizadas em aplicações como as citadas anteriormente, é um grande desafio gerar um modelo de segmentação de veias para a base de dados utilizada, isso porque além do banco de dados não ser muito populoso, técnicas de pré-processamento digital de imagens (PDI) como a aplicação do filtro de Gabor (JINFENG YANG; JINLI YANG; SHI, 2009) foram aplicadas à base e não obtiveram resultados satisfatórios (CARDOSO, 2019).

Para auxiliar na segmentação das veias dos indivíduos pode-se utilizar, por exemplo, inteligência artificial (IA). Avanços em visão computacional e inteligência artificial vem ganhando espaço em diversas áreas, uma das classes de redes neurais que vem se destacando é a das redes neurais convolucionais, (CNNs) (LECUN et al, 1998) com uma ampla gama de problemas de alto nível incluindo classificação de imagem (HASSAN et al, 2020; BODAPATI; VEERANJANEYULU, 2019; XIE; LUONG; HOVY; LE, 2020; SERMANET et al 2013) e detecção de objetos (LI et al, 2019; ZHU et al, 2020; GIRSHICK et al, 2014; REN et al, 2015; LEE et al, 2017).

Algumas CNNs permitem modelar conhecimentos humanos e distinguir diferentes objetos em uma mesma imagem, isso permite a criação de modelos como em (POHLEN et al, 2017) que segmenta imagens capturadas em ruas e avenidas para a construção de um veículo autônomo que consegue diferenciar ruas, pessoas, prédios, calçadas, etc, em outro exemplo como em (MIAO et al, 2018) são utilizadas imagens NIR de dedos para treinar CNN's a reconhecer e segmentar veias.

Neste contexto, este trabalho tem como objetivo construir um modelo de segmentação de veias a partir de uma base de dados de rótulos de imagens de pulso. O modelo foi construído utilizando CNNs de segmentação semântica, o treinamento e teste foram feitos em diferentes cenários, onde a base de dados foi distribuída de diferentes formas a fim de descobrir qual forma se adéqua melhor para o treinamento do modelo.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. Redes Neurais Convolucionais (*Convolutional Neural Network*)

Uma rede neural convolucional (CNN) é um algoritmo de aprendizado profundo, recebe uma imagem de entrada, atribui importância a vários aspectos na imagem e é capaz de diferenciar uns dos outros (KRIZHEVSKY; SUTSKEVER; HINTON, 2017). Detecções de objetos e reconhecimento de rostos são algumas das áreas onde as CNNs são amplamente utilizadas.

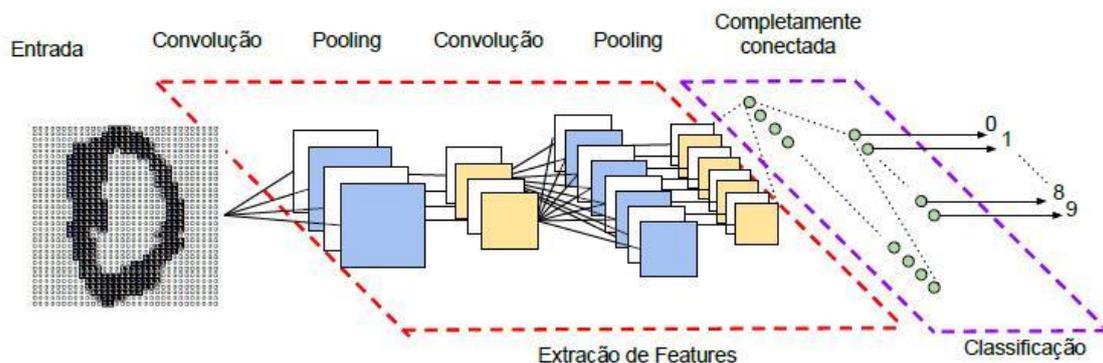
A arquitetura de um CNN é análoga àquela do padrão de conectividade dos neurônios no cérebro humano e foi inspirada na organização do córtex visual. Os neurônios individuais respondem a estímulos apenas em uma região restrita do campo visual conhecido como campo receptivo (ZHANG et al, 2020). Uma coleção de tais campos se sobrepõe para cobrir toda a área visual.

Tecnicamente, modelos de CNN para treinar e testar cada imagem de entrada passam por uma série de camadas de convolução com filtros (*Kernels*), *Pooling*, camadas totalmente conectadas (*FC*) e aplicam geralmente a função *Softmax* para classificar um objeto com valores probabilísticos entre 0 e 1 (ARANDJELOVIC et al, 2016).

A Figura 1 representa de forma simples a arquitetura de uma CNN, é possível observar que as primeiras camadas da rede (marcadas em vermelho) são camadas de Convolução e de *Pooling* e desempenham a função de extrair características da imagem de entrada, tais características são passadas para as camadas de classificação (marcadas em roxo) e através dessas características essas camadas conseguem rotular a imagem de entrada a uma classe de saída (VARGAS et al, 2016).

Um exemplo clássico de CNN é a AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) projetada por Alex Krizhevsky, que em 2012 venceu a competição *Large Scale Visual Recognition Challenge* (ILSVRC), um dos fatores mais importantes para seus resultados foi a profundidade do modelo que era essencial para seu alto desempenho, no entanto, o seu treinamento era caro computacionalmente, mas foi viabilizado devido à utilização de unidades de processamento gráfico (GPUs) durante o treinamento.

**Figura 1 – Exemplo de arquitetura de uma CNN.**



Fonte: (VARGAS et al, 2016)

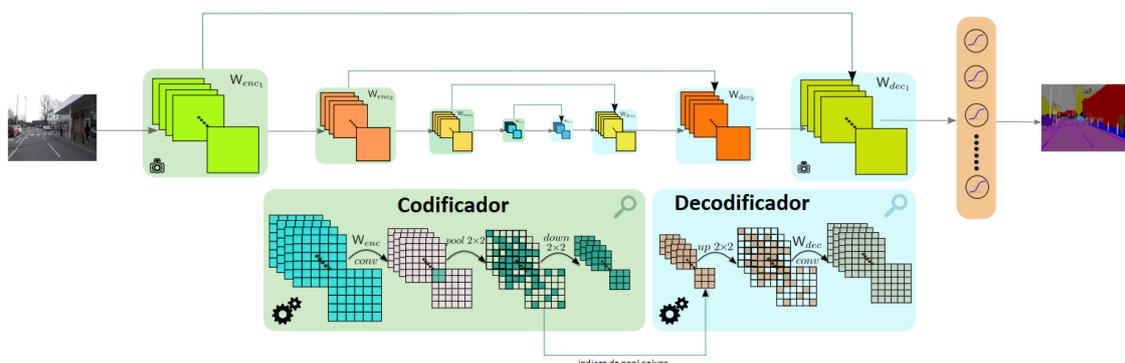
Na atualidade um dos principais problemas que vem ganhando muito destaque é o da segmentação de imagens. Analisando de uma forma simples, a segmentação semântica é uma das tarefas de alto nível que abre o caminho para a compreensão completa de uma cena.

A importância da compreensão de uma cena como um problema central de visão computacional é destacada pelo fato de que, um número crescente de aplicações alimenta a inferência de conhecimento a partir de imagens. Algumas dessas aplicações incluem veículos autônomos (POHLEN et al, 2017; GEIGER; LENZ; URTASUN, 2012; ZHANG; FIDLER; URTASUN, 2016), reconstrução 3D (MALIK; MAYDAN, 1989; LEE; HEBERT; KANADE, 2009), geração de imagem (ISOLA et al, 2017; WANG et al, 2018) e segmentação de veias (JALILIAN; UHL, 2019).

Com a popularidade do aprendizado profundo nos últimos anos, muitos problemas de segmentação semântica estão sendo resolvidos usando arquiteturas profundas, na maioria das vezes por CNNs, que ultrapassam a eficácia de outras abordagens em precisão e eficiência (LONG; SHELHAMER; DARRELL, 2015).

Uma arquitetura geral de segmentação semântica pode ser amplamente considerada como uma rede codificadora, seguida por uma rede decodificadora. Como visto na Figura 2 o codificador é geralmente uma rede de classificação pré-treinada seguida por uma rede de decodificação (BADRINARAYANAN; HANDA; CIPOLLA, 2015). A tarefa do decodificador é projetar semanticamente os recursos discriminativos aprendidos pelo codificador no espaço de píxel para obter uma classificação densa.

**Figura 2 – Exemplo do funcionamento de uma CNN de segmentação semântica.**



Fonte: Badrinarayanan, Vijay e col. (2015)

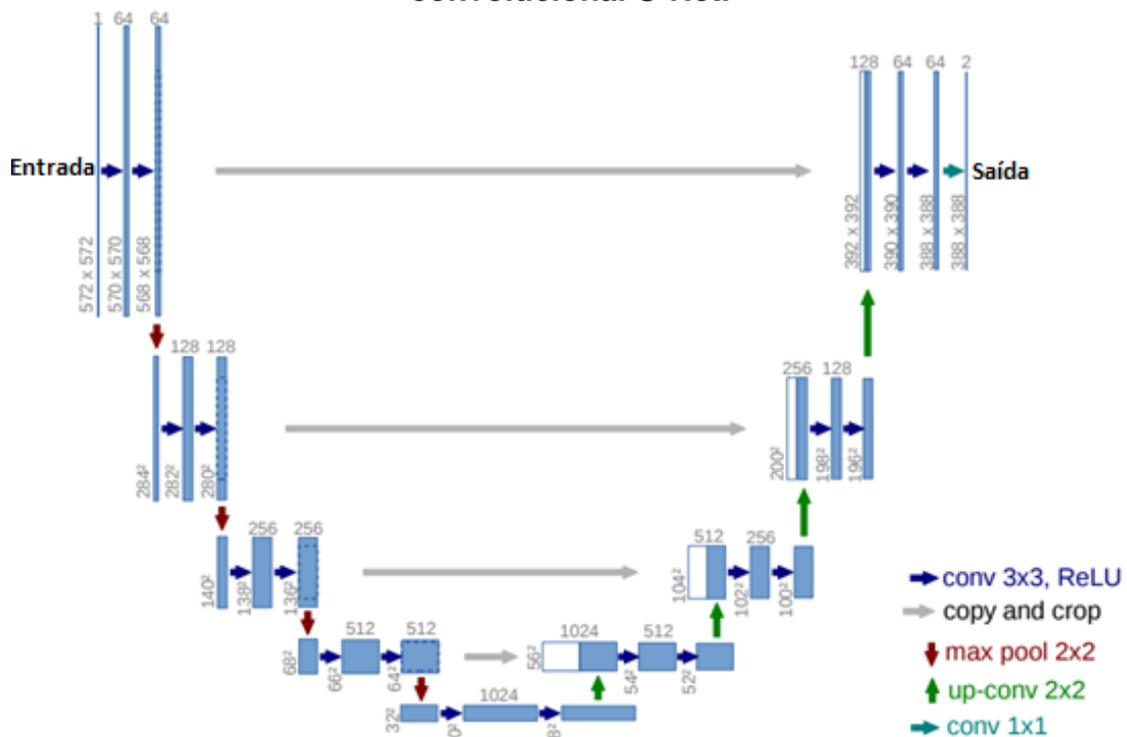
## 2.2. U-Net

A U-Net é uma das mais notáveis CNNs de segmentação de imagens, ela foi desenvolvida pelo departamento de Ciência da Computação da universidade de Freiburg na Alemanha (RONNEBERGER; FISCHER; BROX, 2015). Sua arquitetura foi proposta com o intuito de realizar o treinamento da rede com menos imagens em comparação a outras redes de segmentação e mesmo com poucos dados de treinamento conseguir bons resultados.

A arquitetura da U-Net é representada pela Figura 3, nela é possível observar a mesma arquitetura de codificador-decodificador vista anteriormente, representado a esquerda está o codificador, já representado a direita está o decodificador.

Além de não necessitar de uma quantidade muito grande de dados para seu treinamento a U-Net se destaca por sua velocidade de processamento, uma imagem de dimensões 512 x 512, por exemplo leva menos de um segundo para ser processada pela rede. Tudo isso possibilitou que em 2015 a U-Net vencesse o *Cell Tracking Challenge at ISBI* (RONNEBERGER; FISCHER; BROX, 2015) que é uma importante competição de modelos de segmentação de imagens.

**Figura 3 – Representação da arquitetura da rede neural convolucional U-Net.**



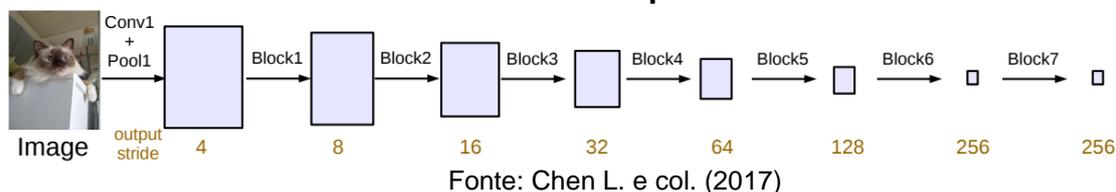
Fonte: Ronneberger, O. e col. (2015)

## 2.3. DeepLabv3

A DeepLabv3 é uma CNN de segmentação semântica de código aberto, desenvolvida pela Google, sua arquitetura está representada pela Figura 4. Como o próprio nome sugere DeepLabv3 provem da evolução da DeepLabv1 e DeepLabv2, nessa evolução mesmo com a remoção de uma camada de pós-processamento *Conditional Random Field* (CRF) que está presente nas duas versões anteriores, o modelo conseguiu resultados muito melhores quando

comparado às outras versões (CHEN et al, 2017). Para o modelo de DeepLabv3 em questão, foi utilizada como rede codificadora a Resnet18.

**Figura 4 – Representação da arquitetura da rede neural convolucional DeepLabv3.**

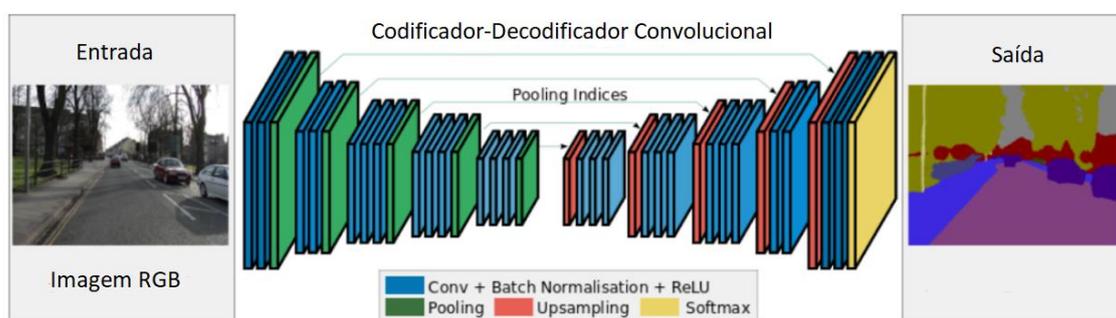


## 2.4. SegNet

A SegNet foi desenvolvida pelo grupo de pesquisa *Computer Vision and Robotics Group* da universidade de Cambridge na Inglaterra sua arquitetura é representada pela Figura 5, o codificador da rede é topologicamente idêntica às 13 camadas convolucionais da rede VGG16 (BADRINARAYANAN; KENDALL; CIPOLLA, 2017).

A SegNet se destaca pela maneira como o decodificador aumenta a amostragem da saída do codificador, que envolve o armazenamento dos índices de agrupamento máximo usados na camada de agrupamento. Isso proporciona um desempenho razoavelmente bom e economiza espaço.

**Figura 5 – Representação da arquitetura da rede neural convolucional SegNet.**



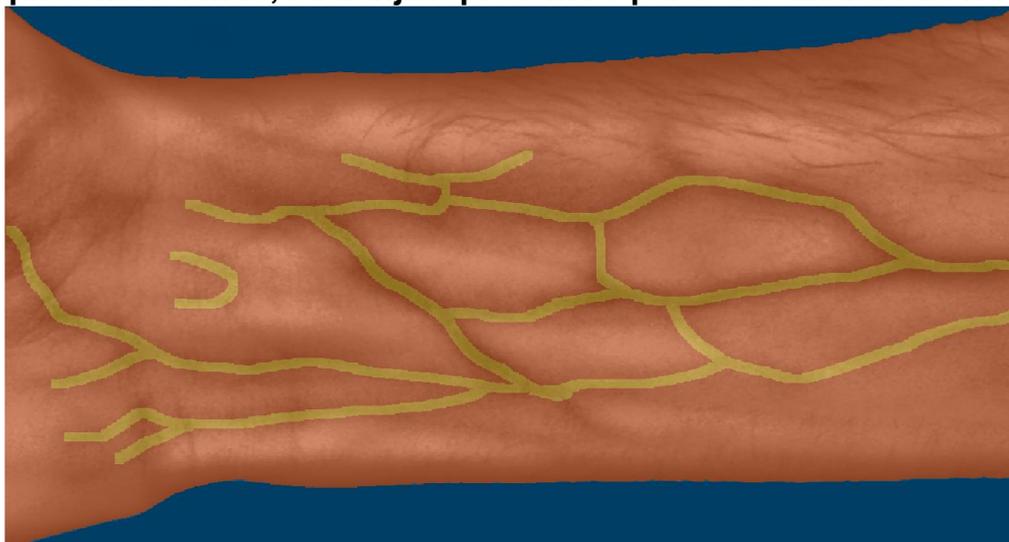
## 2.5. Anotação de imagem (*Image Annotation*)

A anotação no aprendizado de máquina é essencialmente o processo de rotular dados em textos (DONG et al, 2020), vídeos (VENDRIG; WORRING, 2003) ou imagens (CARNEIRO et al, 2007). Os rótulos são geralmente predeterminados por um engenheiro ou cientista de dados, e são escolhidos para fornecer as informações necessárias para que o modelo de visão computacional possa identificar objetos que estão representados em uma imagem (RUSSELL et al, 2007).

O algoritmo então recebe tais dados anotados como entrada para treinar e posteriormente reconhecer padrões semelhantes quando apresentado um novo dado de entrada. Dependendo da natureza do projeto, existem diferentes formas de anotação como, por exemplo anotação de ponto (TONG et al, 2012), caixas delimitadoras (XIA et al, 2018), segmentação semântica (PANG et al, 2019), entre outros.

A segmentação semântica tem como objetivo separar uma imagem em várias seções e classificar cada píxel em cada segmento para um rótulo de classe correspondente do que ele representa, ou seja, pedestre, carro, poste de luz, isso é representado pela Figura 6, onde cada cor representa uma classe. Isso dá ao algoritmo uma compreensão abrangente de cada píxel em uma imagem. É comumente usada para detecção e localização de um objeto específico, isso faz com que tal método seja o mais adequado para a segmentação das veias no trabalho em questão.

**Figura 6 – Exemplo de segmentação semântica onde a cor azul representa o fundo, o laranja representa o pulso e o amarelo as veias.**



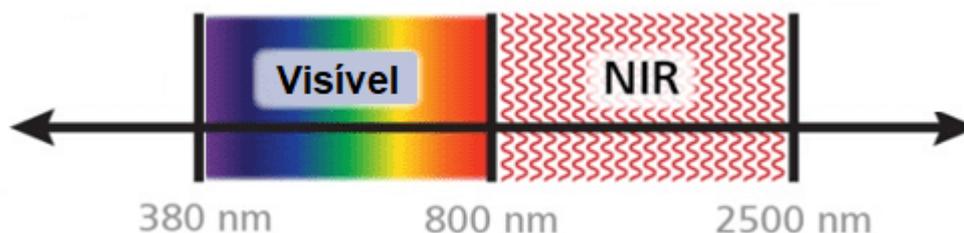
Fonte: Próprio autor.

### 3. MATERIAIS E MÉTODOS

#### 3.1. Estação de Aquisição

Para a realização dessa pesquisa, serão utilizadas imagens providas de uma estação de trabalho JAI AD-080E, que proporciona imagens multiespectrais. Tal câmera trabalha com o canal visível que capta as faixas de RGB (vermelho, verde e azul) (MUDA et al, 2012) e o canal infravermelho próximo (KLARE; JAIN, 2010; HUANG et al, 2014) como representado na Figura 7.

**Figura 7 – Representa as regiões do espectro eletromagnético do visível e do NIR respectivamente.**



Fonte: Próprio autor.

É disponibilizado pela fabricante o software *JAI Control Tool* para controlar a câmera no momento da aquisição, são disponibilizadas opções de fotos no espectro visível ou multiespectral. Todas as informações da câmera estão disponíveis na página da fabricante disponível em (JAI, 2009).

Para o uso da câmera foi construída uma estação de aquisição uma estação de aquisição de imagens multiespectrais, nomeada de SAIM (Sistema de Aquisição de Imagens Multiespectrais) e representada pela Figura 8. O SAIM consiste em um sistema totalmente fechado para não permitir a entrada de luz, nele existem somente duas entradas, uma superior onde é feito o posicionamento da câmera e das lâmpadas e outra frontal onde é inserido o objeto para realizar a aquisição das imagens. A parte superior consiste em uma tampa removível que também é regulável para ser possível aproximar a câmera do objeto. O SAIM foi construído inicialmente para trabalhar com a câmera JAI AD080GE, tendo a possibilidade de se utilizar outras câmeras.

**Figura 8 – Imagem real do SAIM.**



Fonte: Próprio autor.

### **3.2. Base de Dados**

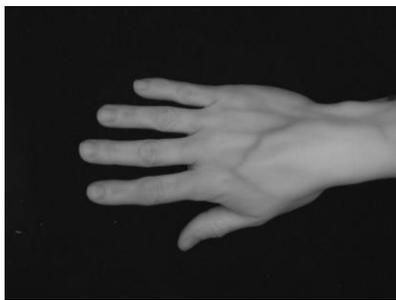
Como estudo de caso, foi gerada uma base de dados para validar o modelo proposto, esta foi toda coletada com o SAIM. Tal base foi projetada para ser objeto de pesquisa deste e de outros projetos envolvendo por exemplo biometria por características da mão, portanto além de possuir imagens no espectro multiespectral, também possui imagens no espectro visível. Futuramente tem-se a intenção de torná-la pública.

A estrutura da base consiste em: para cada indivíduo será adquirido amostras de ambas as mãos; no espectro visível e no espectro NIR; para cada mão será obtido cinco amostras de cada uma das posições marcadas na Tabela 1. Na Figura 9 há uma representação de quais imagens foram adquiridas para a base.

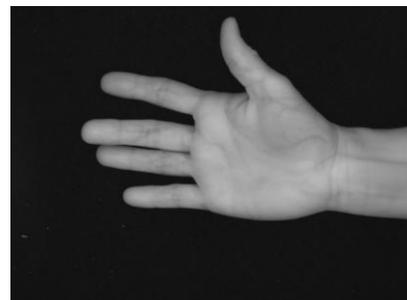
**Tabela 1 - Posições das fotos a serem adquiridas das mãos.**

	<b>Dorso</b>	<b>Palma</b>	<b>Pulso</b>
Câmera Afastada	x	x	
Câmera aproximada	x	x	x

**Figura 9 - Representação das posições da mão que serão obtidas as imagens. A) Dorso afastado. B) Palma afastada. C) Dorso aproximado. D) Palma aproximada. E) Pulso.**



**A**



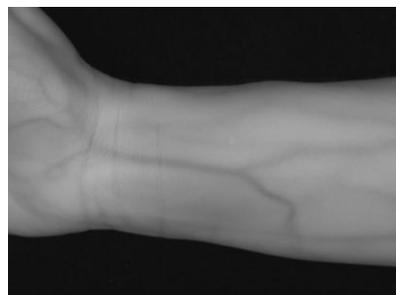
**B**



**C**



**D**



**E**

Fonte: Banco de imagens SAIM.

No total são 100 amostras de cada indivíduo, para a posições de mão onde a camera se encontra afastada tem-se um total de 20 amostras de palma

e de dorso para cada uma das mãos (direita e esquerda), separadas em 10 no espectro visível e 10 em NIR, contabilizando um total de 40 imagens para ambas as mãos. As 60 restantes foram obtidas com a câmera aproximada (palma, pulso e dorso), são 15 no visível e 15 no NIR para cada uma das mãos. A Tabela 2 ilustra a organização da base de dados, nela está representada a nomenclatura das imagens dentro da organização das pastas de acordo com cada posição de mão.

**Tabela 2 - Indexação dos arquivos das fotos. Como serão nomeados e armazenados na pasta. L/R: *Left/Right* (Esquerda/Direita). V/R: *Visible/NIR* (Visível/NIR). H/L: *High/Low* (Afastada/Aproximada). BH/P/W: *Back of the Hand/Palm/Wrist* (Dorso/Palma/Pulso).**

Nº do Indivíduo (pasta)					
XXX					
Nº da Foto	Lado	Espectro	Altura	Região de Interesse (ROI)	Nº da Amostra da Foto
XX	L/R	V/R	H/L	BH/P/W	X

Com o intuito de melhorar a qualidade dos dados, algumas amostras foram descartadas para este estudo e no total foram escolhidas as amostras de 93 indivíduos, totalizando 9300 arquivos. Destes 9300, foram utilizadas somente 465, referentes as amostrar do pulso direito no espectro NIR.

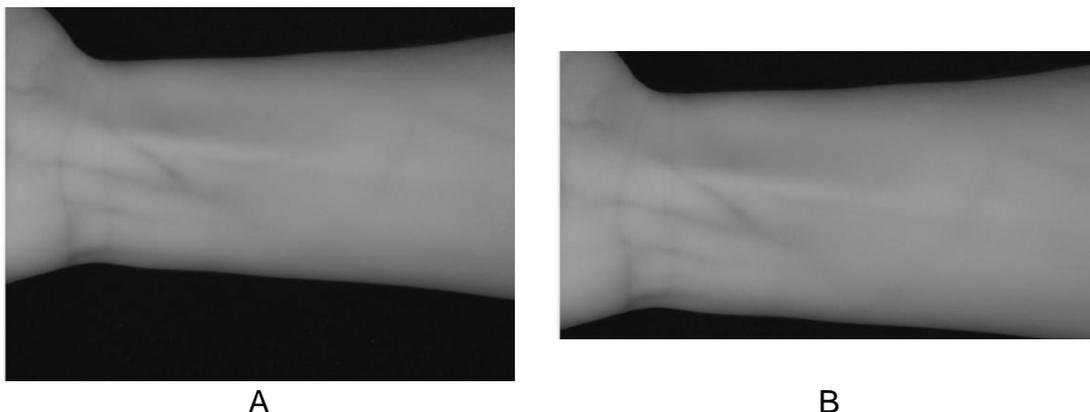
### 3.3. Pré-processamento Digital

O pré-processamento das imagens foi dividido em duas etapas, a primeira tem como objetivo diminuir o máximo da dimensionalidade das imagens com o mínimo de perda de dados relevantes e retirar o máximo de ruídos das imagens.

Para realizar o redimensionamento das imagens, foi utilizada primeiramente uma transformação das imagens visíveis para escala de cinza, em seguida foi feita uma limiarização, depois calculada a soma de todas as submatrizes da imagem e escolhida a que possui maior somatório de acordo com uma janela pré-definida, que corresponde às novas dimensões da imagem, por fim a imagem original é recortada na região de interesse encontrada. O resultado desta primeira etapa pode ser observado na Figura 10.

A primeira etapa do algoritmo de pré-processamento foi desenvolvido desta forma, pois por padrão todas as imagens possuem um fundo que se aproxima do preto, a região do pulso conseqüentemente se destaca em relação ao fundo, fazendo com que ao se aplicar a limiarização, a região onde se encontra o pulso fique totalmente branca e a região do fundo fique totalmente preta, o que permite a aplicação do algoritmo descrito acima, isso faz com que seja possível fazer um corte preciso na imagem eliminando boa parte do fundo.

**Figura 10 - Representação das imagens A) Antes e B) Depois da primeira etapa do pré-processamento.**

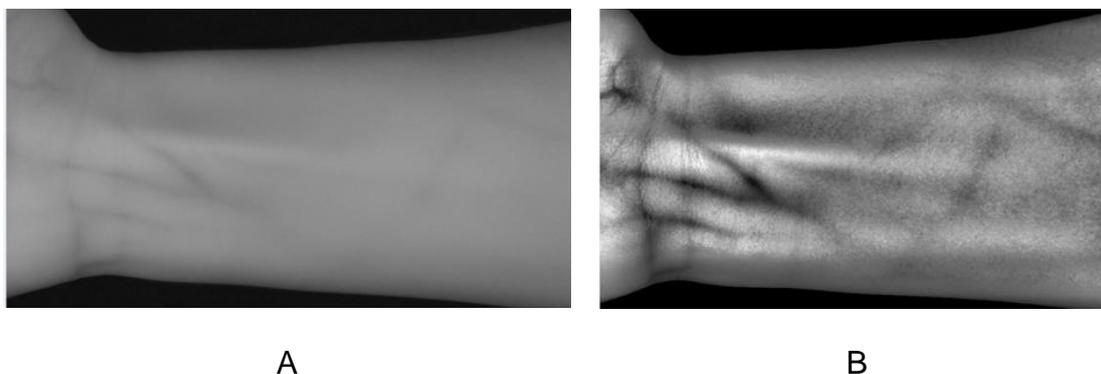


Fonte: Próprio autor.

A segunda etapa tem como objetivo evidenciar as veias baseado no modelo de (MIAO et al, 2018) presentes na imagem para auxiliar no processo de anotação das imagens.

Para evidências as veias na imagem, primeiramente foi aplicado um filtro de médias para minimizar os ruídos, posteriormente foi aplicada a equalização do histograma adaptativo da imagem, em seguida foi aplicado novamente um filtro de médias para retirar alguns ruídos gerados pela equalização, depois é feita uma binarização para se criar uma máscara e por fim é aplicada a máscara na imagem resultante da última aplicação do filtro de médias. A Figura 11 tem a representação da diferença da imagem com e sem o pré-processamento.

**Figura 11 - Representação das imagens A) Antes e B) Depois da segunda etapa do pré-processamento.**

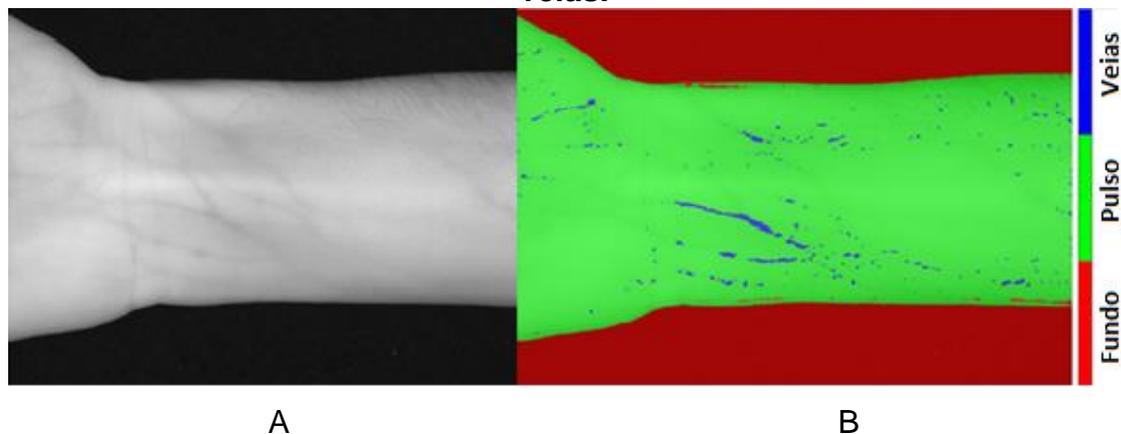


Fonte: Próprio autor.

### **3.4. Construção do Modelo e Simulação**

Para construção do modelo, os estudos foram abordados na prática e a base foi testada. Com um pequeno protótipo baseado no modelo de segmentação de (JALILIAN; UHL, 2019) foram feitos alguns testes de segmentação usando as fotos do dorso direito com a câmera afastada, tais testes são representados pela Figura 12.

**Figura 12 - Segmentação de uma imagem do indivíduo 83: A) Imagem original presente no conjunto de treinamento, B) Imagem retornada pela CNN, com a cor vermelha para o fundo, verde para o pulso e azul para veias.**



Fonte: Próprio autor.

Para o trabalho em questão foram utilizadas somente as imagens do pulso direito retiradas da base de dados original, isso porque a região do pulso contém menos ruídos quando comparada com outras regiões, o dorso por exemplo contém mais pelos que a região do pulso, além disso a região do pulso não possui linhas como as presentes na palma da mão.

A base de dados de pulso foi separada de três formas diferentes para dados de treinamento e teste. Para o modelo A os primeiros 56 indivíduos da base foram selecionados para treinamento e os últimos 37 para teste, posteriormente para o modelo B foram separadas três imagens de cada indivíduo para treinamento e duas para teste e por fim para o modelo C as amostras de treinamento e teste foram escolhidas aleatoriamente sendo 60% da base para treinamento e 40% para teste.

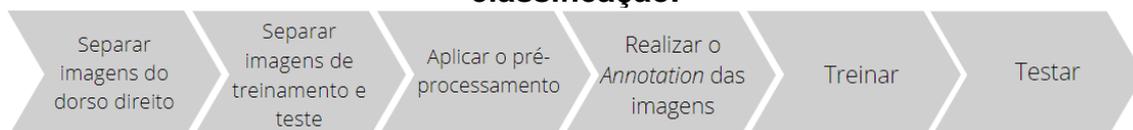
Para a identificação do modelo de CNN ideal para segmentação das veias foram treinados três modelos de CNN de segmentação (DeepLabv3, U-Net, Segnet), foi utilizado o modo de treinamento A, então o resultado das redes foi comparado para que fosse selecionado o melhor modelo, em seguida a CNN selecionada foi treinada para os outros dois modelos de treinamento (B e C). Foram seguidos os seguintes passos:

1. Separação das imagens somente da região do dorso direito;
2. Aplicar o pré-processamento nas imagens;
3. Realizar a anotação de todas as imagens;
4. Separar imagens e seus respectivos rótulos para o treinamento e para o teste de acordo com o modelo de treinamento A;
5. Treinar diferentes CNNs com diferentes topologias;
6. Analisar os resultados e escolher qual CNN obteve a melhor acurácia;
7. Treinar a CNN escolhida com diferentes configurações de dados de treinamento e teste.
8. Segmentar algumas das imagens de teste para obter resultados visuais da segmentação;

Todos os passos descritos anteriormente são representados pela Figura 13. O *software* MATLAB foi utilizado para construção e simulação do modelo e

todos os parâmetros utilizados para o treinamento e teste das CNN's foram os padrões indicados pelo próprio MATLAB. A fase de anotação das imagens demandou muito tempo, conseqüentemente não restou tempo para a realização de análises adicionais como, por exemplo treinar os modelos com diferentes tipos de parametrização.

**Figura 13 - Fluxograma representando o modelo proposto para classificação.**



Fonte: Próprio autor.

#### 4. RESULTADOS

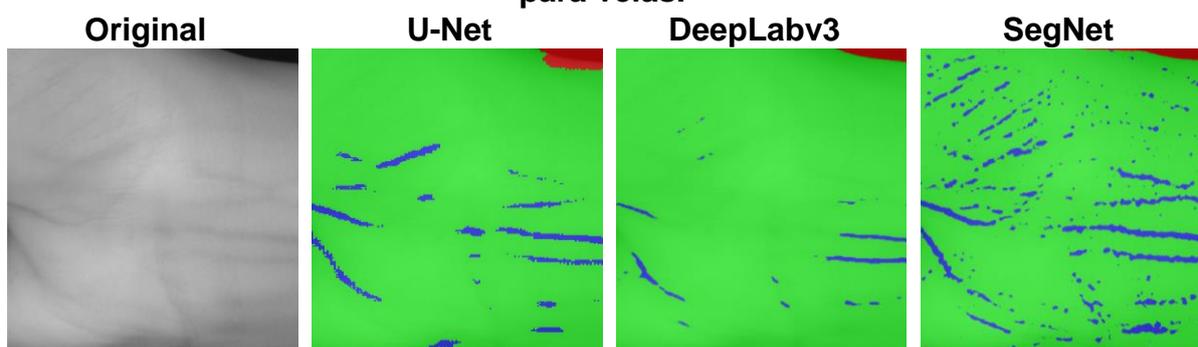
Após o treinamento do modelo e a simulação foram obtidos os seguintes resultados apresentados pela Tabela 3, quando comparados os modelos de CNN U-Net, DeepLabv3 e Segnet. O treinamento foi feito com o modelo A realizado com os primeiros 56 indivíduos para treinamento e os últimos 37 para teste.

Na Figura 14 é possível observar as saídas de cada uma das redes, nela é possível observar nitidamente que, a Segnet identifica muito mais regiões com veias em comparação com as outras redes, além disso, é possível observar que todas as três redes conseguiram anotar regiões onde as veias estão com maior destaque na imagem de entrada.

**Tabela 3 – Desempenho das redes (%), com treinado de acordo com o modelo de treinamento A.**

<b>LABELS</b>	<b>U-NET</b>	<b>DEEPLABV3</b>	<b>SEGNET</b>
FUNDO	83.99%	98.23%	98.65%
PULSO	98.21%	98.62%	82.48%
VEIAS	12.08%	42.86%	86.52%
MÉDIA	64.76%	79.90%	89.31%

**Figura 14 – Representação a saída de cada CNN treinada para uma imagem de teste, sendo vermelho para Fundo, verde para o pulso e azul para veias.**



Fonte: Próprio autor.

Foi possível observar que as redes U-Net e DeepLabv3 tiveram acurácia para o reconhecimento das veias abaixo de 50%, isso porque a base de imagens anotadas é muito pequena o que dificulta o aprendizado das redes, além disso, a rede DeepLabv3 se mostrou eficiente para o reconhecimento do fundo e do pulso com acurácia em ambos acima dos 98%.

Foi possível verificar que na comparação entre as redes, a SegNet teve o melhor resultado para o reconhecimento das veias, então para tal rede foram realizados os outros testes quanto a forma de organização dos dados de treinamento e teste e foram obtidos os resultados demonstrados na Tabela 4.

**Tabela 4 – Desempenho da rede (%), com treinado de acordo com os modelos de treinamento A, B e C.**

<b>SEGET</b>			
<b>LABELS</b>	<b>A</b>	<b>B</b>	<b>C</b>
FUNDO	98.65%	98.07%	98.65%
PULSO	82.48%	85.10%	85.20%
VEIAS	86.52%	85.13%	84.05%
MÉDIA	89.31%	89.43%	89.30%

Através destes testes foi constatado que, o modelo de treinamento A teve melhores resultados quanto ao reconhecimento das veias em comparação aos outros modelos, isso porque os modelos B e C utilizam imagens de um mesmo indivíduo para treinamento e teste, então se alguns erros mesmo que pequenos tenham ocorrido no decorrer das anotações, impactam diretamente na validação dos dados de teste.

Tais resultados com uma base de dados maior incluído outras regiões como pulso e palma podem fazer com que as redes U-Net e DeepLabv3 alcancem resultados melhores que a SegNet para a segmentação das veias.

## **5. CONCLUSÃO**

A segmentação de imagens é uma parte muito importante em diversas aplicações. Nesta fase, vários métodos de segmentação são explorados para obter a maior acurácia no processo de segmentação de uma imagem, e para que tais algoritmos consigam um bom resultado, os dados de entrada devem ser gerados da forma mais precisa possível, utilizando a técnica de anotação que mais se adéqua ao problema. Nesse trabalho foram explorados modelos de CNNs para segmentação de imagens no intuito de segmentar as veias em imagens NIR de pulso.

Foi possível observar que dentre os modelos escolhidos a SegNet foi o que mais se destacou para a segmentação das veias, além disso, o modelo de treinamento que obteve os melhores resultados para a rede escolhida (Segnet), foi o modelo onde foram utilizados indivíduos distintos para treinamento e teste (modelo A).

É possível que os modelos de treinamento B e C tiveram valores de acurácia para as veias mais baixos que o modelo A, porque nesses dois modelos um mesmo indivíduo tem imagens para treinamento e teste, considerando que a base foi gerada manualmente erros humanos são passíveis de acontecer, isso

quer dizer que podem haver pequenas divergências em relação à anotação das imagens de treinamento e teste em um mesmo indivíduo, fazendo com que o algoritmo tenha maior dificuldade de rotular os pixels da imagem de maneira correta.

Não foi possível chegar a uma acurácia de 100% de acerto para veias em nenhum dos casos, isso porque, em geral as CNN's exigem que a base de dados para treinamento seja grande e com dados diversificados, no entanto, no cenário em questão podemos considerar que a base de dados usada nos treinamentos não é muito grande e que os dados não estão diversificados, pois são somente de imagens do pulso.

Neste contexto, pretende-se, também em trabalhos futuros, aumentar a base de dados para o treinamento das CNN's e diversificar a mesma com imagens de outras regiões como dorso e palma.

## REFERÊNCIAS

- [1] WANG, L.; LEEDHAM, G. **Near- and Far- Infrared Imaging for Vein Pattern Biometrics**. Em: 2006 IEEE International Conference on Video and Signal Based Surveillance, 2006, p. 52.
- [2] JENSEN, J. R. **Introductory Digital Image Processing: A Remote Sensing Perspective**. Univ. of South Carolina, Columbus, 1986.
- [3] KLARE, B.; JAIN A. K. **Heterogeneous face recognition: Matching NIR to visible light images**. Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, p. 1513-1516, 2010.
- [4] CHENG-TANG, P.; MARK, F. D.; CHUNG-KUN, Y.; SHAO-YU. W.; YOW-LING, S. (2019). **Vein Pattern Locating Technology for Cannulation: A Review of the Low-Cost Vein Finder Prototypes Utilizing near Infrared (NIR) Light to Improve Peripheral Subcutaneous Vein Selection for Phlebotomy**. Sensors, 19(16), 3573–.doi:10.3390/s19163573.
- [5] AGARWAL, A.; MAHESHWARI, S.; YADAV, G. (2014). **A review on vein biometric recognition using geometric pattern matching techniques**. 2014 Conference on IT in Business, Industry and Government (CSIBIG). doi:10.1109/csibig.2014.7056935.
- [6] SARKAR, I. et al. **Palm vein authentication system: a review**. International Journal of Control and Automation v. 3, n. 1, p. 27-34, 2010.
- [7] KUMARI A., ALANKAR B., GROVER, J. **Feature Level Fusion of Multispectral Palmprint**. International Journal of Computer Applications, New York: Foundation of Computer Science (FCS), v. 144, n. 3, p. 41- 6, 2016.
- [8] BHOKARE, R. et al. **Multispectral Palm Image Fusion: A Critical Review**. International Journal of Advanced Research in Computer Engineering & Technology, v. 2, n. 6, p. 2159-2164, 2013.
- [9] YANG, JINFENG; YANG, JINLI; SHI, Y. (2009). **Finger-vein segmentation based on multi-channel even-symmetric Gabor filters**. 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems. doi:10.1109/icicisys.2009.5357631.
- [10] CARDOSO, M. **Pré-processamento de imagens das mãos para retirada de ruídos e diminuição de dimensionalidade**. V Congresso de Ciência e Tecnologia, Goiânia, 2019.
- [11] LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. **Gradient-based learning applied to document recognition**. Proc. IEEE, 1998.

- [12] HASSAN, S.A.; SAYED, M.S.; ABDALLA, M.I. et al. **Breast cancer masses classification using deep convolutional neural networks and transfer learning**. *Multimed Tools Appl* 79, 30735–30768 (2020). <https://doi.org/10.1007/s11042-020-09518-w>.
- [13] BODAPATI, J.; VEERANJANEYULU, N. **Feature Extraction and Classification Using Deep Convolutional Neural Networks**. *Journal of Cyber Security and Mobility*, [S. l.], p. 261-276, 25 jan. 2019.
- [14] XIE, Q.; LUONG, M; HOVY, E; LE, Q. V. **Self-training with Noisy Student improves ImageNet classification**. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 10687-10698 p.
- [15] SERMANET, P.; EIGEN, D.; ZHANG X.; MATHIEU, M.; FERGUS, R.; LECUN, Y. **Overfeat: Integrated recognition, localization and detection using convolutional networks**. arXiv:1312.6229, 2013.
- [16] LI, Z.; DONG, M.; WEN, S.; HU, X.; ZHOU, P.; ZENG, Z. (2019). **CLU-CNNs: Object Detection for Medical Images**. *Neurocomputing*. doi:10.1016/j.neucom.2019.04.028.
- [17] ZHU, H.; YAN, X.; TANG, H.; CHANG, Y.; LI, E. B.; YUAN, F. X. **Moving object detection with Deep CNNs**. *IEEE Access* 2020. 1 p. doi:10.1109/access.2020.2972562.
- [18] GIRSHICK, R.; DONAHUE, J.; DARRELL, T.; MALIK J. **Rich feature hierarchies for accurate object detection and semantic segmentation**. *CVPR*, 2014.
- [19] REN, S.; HE, K.; GIRSHICK, R.; SUN, J. **Faster r-cnn: Towards realtime object detection with region proposal networks**. *NIPS*, 2015.
- [20] LEE, J.; WANG, J., CRANDALL, D.; SABANOVIC, S.; FOX, G. (2017). **Real- Time, Cloud-Based Object Detection for Unmanned Aerial Vehicles**. 2017 First IEEE International Conference on Robotic Computing (IRC). doi:10.1109/irc.2017.77.
- [21] POHLEN, T.; HERMANS, A.; MATHIAS, M.; LEIBE, B. **Fullresolution residual networks for semantic segmentation in street scenes**. *CVPR*, 2017. 1, 2.
- [22] MIAO, Y.; SONG, D.; SHI, W.; YANG, H.; LI, Y.; JIANG, Z.; GU, W. (2018). **Application of the CLAHE Algorithm Based on Optimized Bilinear Interpolation in Near Infrared Vein Image Enhancement**. *Em: 2nd International Conference on Computer Science and Application Engineering - CSAE '18*. doi:10.1145/3207677.

- [23] KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. (2017). **ImageNet classification with deep convolutional neural networks**. *Communications of the ACM*, 60(6), 84–90. doi:10.1145/3065386.
- [24] ZHANG, T.; ZENG, Y.; PAN, R.; SHI, M.; LU, E. (2020). **Brain-Inspired Active Learning Architecture for Procedural Knowledge Understanding Based on Human-Robot Interaction**. *Cognitive Computation*. doi:10.1007/s12559-020-09753-1.
- [25] ARANDJELOVIC, R.; GRONAT, P.; TORII, A.; PAJDLA, T.; SIVIC, J. **NetVLAD: CNN Architecture for Weakly Supervised Place Recognition**. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5297-5307.
- [26] VARGAS, A.; PAES, A.; VASCONCELOS, C. **Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres**. In: *Proceedings of the XXIX Conference on Graphics, Patterns and Images*. 2016. p. 1-4.
- [27] GEIGER, A.; LENZ, P.; URTASUN, R. **Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite**. Em: *CVPR*, 2012. 1.
- [28] ZHANG, Z.; FIDLER, S.; URTASUN, R. **Instance-level segmentation for autonomous driving with deep densely connected mrfs**. Em: *CVPR*, 2016. 1.
- [29] MALIK, J.; MAYDAN, D. E. **Recovering three-dimensional shape from a single image of curved objects**. *T-PAMI*, 11(6):555–566, 1989. 1.
- [30] LEE, D. C.; HEBERT, M.; KANADE, T. **Geometric reasoning for single image structure recovery**. Em: *CVPR*, pages 2136– 2143, 2009. 1.
- [31] ISOLA, P.; ZHU, J. Y.; ZHOU, T.; EFROS, A. A. **Image-to-image translation with conditional adversarial networks**. Em: *CVPR*, 2017. 1.
- [32] WANG, T. C.; LIU, M.Y.; ZHU, J.Y.; TAO, A.; KAUTZ, J.; CATANZARO, B. **High-resolution image synthesis and semantic manipulation with conditional gans**. Em: *CVPR*, 2018. 1.
- [33] JALILIAN, E.; UHL A. **Enhanced Segmentation-CNN based Finger-Vein Recognition by Joint Training with Automatically Generated and Manual Labels**. *2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, Hyderabad, India, 2019, pp. 1-8, doi: 10.1109/ISBA.2019.8778522.
- [34] LONG, J.; SHELHAMER, E.; DARRELL, T. **Fully Convolutional Networks for Semantic Segmentation**. Em: *CVPR*, 2015. 1, 2.

- [35] BADRINARAYANAN, V.; HANDA, A.; CIPOLLA, R. **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling**. CoRR abs/1505.07293 (2015).
- [36] RONNEBERGER, O.; FISCHER, P.; BROX, T. (2015). **U-Net: Convolutional Networks for Biomedical Image Segmentation**. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 234–241. doi:10.1007/978-3-319-24574-4\_28.
- [37] CHEN, L.; PAPANDREOU, G.; SCHROFF, F.; ADAM H. (2017). **Rethinking Atrous Convolution for Semantic Image Segmentation**. CoRR, abs/1706.05587.
- [38] BADRINARAYANAN V.; KENDALL A.; CIPOLLA R. (2017). **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (), 1–1. doi:10.1109/TPAMI.2016.2644615.
- [39] DONG, H.; WANG, W.; HUANG K.; COENEN F. (2020) **Automated Social Text Annotation With Joint Multilabel Attention Networks**. Em: IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2020.3002798.
- [40] VENDRIG, J.; WORRING, M. **Interactive adaptive movie annotation**. Em: IEEE MultiMedia, vol. 10, no. 3, pp. 30-37, July-Sept. 2003, doi: 10.1109/MMUL.2003.1218254.
- [41] CARNEIRO, G.; CHAN, A. B.; MORENO P. J.; VASCONCELOS, N. **Supervised Learning of Semantic Classes for Image Annotation and Retrieval**. Em IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 3, pp. 394-410, March 2007, doi: 10.1109/TPAMI.2007.61.
- [42] RUSSELL, B. C.; TORRALBA, A.; MURPHY, K. P.; FREEMAN, W. T. (2007). **LabelMe: A Database and Web-Based Tool for Image Annotation**. International Journal of Computer Vision, 77(1-3), 157–173. doi:10.1007/s11263-007-0090-8.
- [43] TONG, Y.; LIU, X.; WHEELER, F. W.; TU, P. H. (2012). **Semi-supervised facial landmark annotation**. Computer Vision and Image Understanding, 116(8), 922–935. doi:10.1016/j.cviu.2012.03.008.
- [44] XIA, G.; BAI, X.; DING, J.; ZHU, Z.; BELONGIE, S.; LUO, J.; DATCU, M.; PELILLO, M.; ZHANG, L. **DOTA: A Large-Scale Dataset for Object Detection in Aerial Images**. Em: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp.3974-3983.

- [46] PANG, Y.; LI, Y.; SHEN, J.; SHAO, L. **Towards Bridging Semantic Gap to Improve Semantic Segmentation.** Em: IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 4230-4239.
- [47] MUDA, M. A.; FOULONNEAU, A.; BIGUE, L.; SUDIBYO, H.; SUDIANA, D. (2012). **Small format optical sensors for measuring vegetation indices in remote sensing applications: A comparative approach.** TENCON 2012 IEEE Region 10 Conference. doi:10.1109/tencon.2012.6412302.
- [48] HUANG, W.; LI, J.; WANG, Q.; CHEN, L. (2015). **Development of a multispectral imaging system for online detection of bruises on apples.** Journal of Food Engineering, 146, 62–71. doi:10.1016/j.jfoodeng.2014.09.002
- [49] JAI. 2009. Fusion Series AD-080GE. **JAI - See the Possibilities.** [Online] JAI, Outubro de 2009. [Citado em: 10 de Dezembro de 2017.] <https://www.jai.com/products/ad-080-ge>.