



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIA E TECNOLOGIA
DEPARTAMENTO DE CIÊNCIAS DA COMPUTAÇÃO
CURSO DE CIÊNCIAS DA COMPUTAÇÃO**

MARCELO BEZERRA GUSMÃO JUNIOR

**APLICANDO MÉTRICAS DE AVALIAÇÃO NÃO-SUPERVISIONADAS PARA
IMAGENS SEGMENTADAS**

**CAMPINA GRANDE - PB
2020**

MARCELO BEZERRA GUSMÃO JUNIOR

**APLICANDO MÉTRICAS DE AVALIAÇÃO NÃO-SUPERVISIONADAS PARA
IMAGENS SEGMENTADAS**

Trabalho de Conclusão de Curso apresentado à
Universidade Estadual da Paraíba, como
requisito parcial à obtenção do título em
Bacharel em Computação.

Área de concentração: Teste de Software.

Orientadora: Profa. Dra. Sabrina de Figueirêdo Souto

**CAMPINA GRANDE - PB
2020**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

G982a Gusmão Junior, Marcelo Bezerra.
Aplicando métricas de avaliação não-supervisionadas para imagens segmentadas [manuscrito] / Marcelo Bezerra Gusmão Junior. - 2020.
42 p. : il. colorido.
Digitado.
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2020.
"Orientação : Profa. Dra. Sabrina de Figueirêdo Souto, Coordenação do Curso de Computação - CCT."
1. Diagnóstico médico. 2. Segmentação de imagem. 3. Aprendizado de máquina. I. Título

21. ed. CDD 600

MARCELO BEZERRA GUSMÃO JUNIOR

**APLICANDO MÉTRICAS DE AVALIAÇÃO NÃO-SUPERVISIONADAS PARA
IMAGENS SEGMENTADAS**

Trabalho de Conclusão de Curso de
Graduação em Ciência da Computação da
Universidade Estadual da Paraíba, como
requisito à obtenção do título de Bacharel em
Ciência da Computação.

Aprovada em 27 de fevereiro de 2020.



Prof.a. Dra. Sabrina de Figueirêdo Souto (DC - UEPB)
Orientador(a)



Prof.a. ra. Kézia de Vasconcelos Oliveira Dantas (DC - UEPB)
Examinador(a)



Prof. Dr. Paulo Eduardo e Silva Barbosa (DC - UEPB)
Examinador(a)

RESUMO

A segmentação de imagens é uma importante etapa de processamento em muitos aplicativos de imagem, vídeo e visão computacional. Pesquisas extensivas foram feitas na criação de muitas abordagens e algoritmos diferentes para a segmentação de imagens, mas ainda é difícil avaliar se um algoritmo produz segmentações mais precisas que outro, seja para uma imagem ou conjunto de imagens em particular, ou de maneira mais geral, para toda uma classe de imagens. Com isso existem os métodos de avaliação para poder validar a imagem segmentada ou o conjunto de imagens. Neste trabalho, buscamos e estudamos os métodos de avaliação para poder validar um conjunto de imagens que são usados para diagnóstico médico. Foram encontrados métodos de avaliação analítica, supervisionado e não supervisionado. Os métodos de avaliação que requerem assistência ao usuário, como avaliação analítica e avaliação supervisionada, são inviáveis em muitas aplicações de visão, por não medir a qualidade da imagem ou não ter uma imagem de referência. A avaliação não supervisionada permite a comparação objetiva de métodos diferentes de segmentação e parametrizações diferentes de um único método, sem a necessidade de comparações visuais humanas ou comparação com uma imagem de referência segmentada manualmente ou pré-processada. Além disso, métodos não supervisionados geram resultados para imagens individuais e imagens cujas características podem não ser conhecidas até o momento da avaliação. Os métodos de avaliação não supervisionado foram aplicados neste trabalho a dois conjuntos de imagens, um de imagens naturais e outro com imagens de feridas. O resultado da pesquisa mostra que por falta de alguns ajustes do código da segmentação, as imagens podem ter uma super segmentação ou uma sub-segmentação. Por fim, conclui-se que o algoritmo de segmentação precisa de um ajuste que possa melhorar o seu resultado tendo uma imagem escura ou clara com entrada do algoritmo.

Palavras-Chave: Diagnóstico Médico, Segmentação de Imagem, Avaliação de Segmentação

ABSTRACT

The image segmentation is an important phase of processing in many apps of image, videos and computational sight. Extensive research were made in the creation of many approaches and different algorithms to the image segmentation, however it is still difficult to evaluate if an algorithm produces more precise segmentations than others, being it an image in specific or a more general way, for a whole image classification. There are methods of evaluation that can validate the segmented image or the set of images. In this project, we search and study the methods of evaluation so we validate a set of images that are used for a medical diagnosis. It was found analytical evaluation methods - supervised and non-supervised. The evaluation methods that requires user assistance as analytical evaluation and supervised evaluation are impracticable on many sight applications, to not be able to measure the quality of the image or not be able to have an image reference. The non-supervised evaluation allows the objective comparison of different methods of segmentation and different parameterizations of one single method, without the need of human visuals comparison or a manual segmented image reference or pre-processed. Besides, non-supervised methods generates results for a individual image and images whose characteristics cannot be known until the evaluation moment. The method of non-supervised evaluation were applied to both sets of images, one of natural images and other of wound images. The research results showed that for lack of some adjustments on the code of segmentation, the images could come with an excessive segmentation or a subsegmentation. Lastly, it is concluded that the segmentation algorithm needs a better adjustment that can improve the results having a dark or clear image with an algorithm entrance.

Keywords: Medical diagnostic, Image Segmentation, Evaluation Segmentation

LISTA DE ILUSTRAÇÕES

Figura 1 - Exemplos de segmentação	12
Figura 2 - Diagrama de avaliação de segmentação supervisionada	14
Figura 3 - Exemplo de avaliação de segmentação não supervisionada.....	15
Figura 4 - Imagem 17, ver Tabela 1 para comparar dados das métricas.	28
Figura 5 - Imagem 11, ver Tabela 1 para análise das métricas.	29
Figura 6 - Imagem 7, ver Tabela 1 para análise das métricas.	29
Figura 7 - Granulado 27. Ver dados das métricas na Tabela 2.....	32
Figura 8 - Granulado 15. Ver dados das métricas na Tabela 2.....	32

LISTA DE TABELAS

Tabela 1 - Resultados do BSDS	26
Tabela 2 - Resultado do Banco de Imagens de Feridas	30

Sumário

1	INTRODUÇÃO	8
1.1	Problema	8
1.2	Solução	9
1.3	Objetivos	10
1.4	Estrutura do Documento	10
2	REFERENCIAL TEÓRICO	11
2.1	Aprendizado de Máquina	11
2.1.1	<i>Segmentação de Imagem</i>	12
2.2	Avaliação de Segmentação	13
2.2.1	<i>Método Analítico</i>	13
2.2.2	<i>Método Supervisionado</i>	14
2.2.3	<i>Método Não Supervisionado</i>	15
2.2.4	<i>Métricas de avaliação não supervisionadas</i>	16
3	MÉTRICAS DE AVALIAÇÃO DE SEGMENTAÇÃO	19
3.1	Revisão Bibliográfica Exploratória	19
3.2	Métricas de Avaliação Supervisionadas	19
3.3	Métricas de Avaliação Não-Supervisionadas	21
4	APLICANDO MÉTRICAS DE AVALIAÇÃO NÃO SUPERVISIONADAS	25
5	ESTUDO DE CASO	26
5.1	Berkeley Segmentation Data Set (BSDS)	26
5.1.1	<i>Resultados</i>	26
5.1.2	<i>Discussão</i>	28
5.2	Banco de Imagens de Feridas	30
5.2.1	<i>Resultados</i>	30
5.2.2	<i>Discussão</i>	31
6	CONCLUSÃO	33
	REFERÊNCIAS	34
	APÊNDICE	39

1 INTRODUÇÃO

Inteligência Artificial (IA) é o domínio da Ciência e Engenharia relacionado à teoria e prática de desenvolvimento de sistemas que exibem as características que associamos a inteligência no comportamento humano, como percepção, processamento de linguagem natural, resolução e planejamento de problemas, aprendizado e adaptação, que atua no meio ambiente. Seu principal objetivo científico é entender os princípios que permitem um comportamento inteligente em humanos, animais e agentes artificiais (TECUCI, 2012). A IA é um campo interdisciplinar muito amplo que tem raízes e se cruza com muitos domínios, não apenas as disciplinas de computação, mas também matemática, linguística, psicologia, neurociência, engenharia mecânica, estatística, economia, teoria de controle e cibernética, filosofia e muitos outros.

O progresso nas áreas da IA levou a um interesse renovado no desenvolvimento de agentes que integram múltiplas funções cognitivas. Isso, por sua vez, levou à compreensão de que várias abordagens e métodos desenvolvidos nos subcampos isolados da IA, como o Aprendizado de máquina.

De acordo com Santos (2005), o Aprendizado de máquina nada mais é que um aprendizado por experiência, que conforme a tarefa é executada, o problema aprende a melhor maneira de se resolver, além de estruturar o conhecimento existente, para levar a um entendimento do aprendizado. Vendo isso podemos chegar à segmentação de imagem, e o Aprendizado de máquina pode favorecer esta área, podendo segmentar a área desejada de uma imagem.

A segmentação de imagem é o processo de decompor uma imagem em várias regiões. Após a segmentação, representações mais significativas podem ser adquiridas para posterior compreensão e análise da imagem. Como resultado, a segmentação serve como um passo fundamental em muitas aplicações de visão computacional (GONG et., 2015).

1.1 Problema

A principal dificuldade para o uso da segmentação é como validar a sua imagem segmentada, visto que dependendo do caso que esteja sendo usado, o grau de validação pode variar. Por exemplo, para análise médica esse fator pode ser crítico, tornando necessário atingir um alto grau de qualidade na segmentação, para reduzir o risco de um diagnóstico médico errado.

Nesse contexto, alguns fatores podem impactar no resultado da segmentação. Muitas vezes pode ocorrer da imagem não está bem iluminada, a mesma pode conter algum tipo de ruído, ou a resolução da imagem pode não estar apropriada, tudo isso pode ter um grande impacto no resultado. Um problema-chave na segmentação é o de se dividir em poucas regiões (sub-segmentação) ou em muitas regiões (super-segmentação). Isso é facilmente demonstrado pelos meus primeiros resultados, como mostrado abaixo (Marco's Blog, acesso em 29/02/2020).

1.2 Solução

Com o intuito de resolver esse problema, várias abordagens foram propostas ZHANG et al. 2004 propõem um método de avaliação objetivo baseado na entropia da imagem, YIN et al., 2017 desenvolveu um método de avaliação através da entropia difusa, CHEN e WANG, 2004 uso de diferença de cor visível em um novo esquema de avaliação quantitativa para segmentação de imagem em cores. Dentre elas, temos abordagens supervisionadas e não supervisionadas. No primeiro caso, a abordagem precisa de uma imagem de referência para poder avaliar se a imagem foi bem segmentada, por exemplo, a abordagem 1 proposta por PENG et al. (2017), onde o autor cria um banco de imagens segmentadas por humanos e usa essas imagens segmentadas como referência para avaliar a segmentação das imagens feitas pelo algoritmo. Já abordagem não supervisionada não precisa de uma imagem de referência para avaliar a imagem segmentada. Nessas, são aplicadas métricas em relação a atributos da imagem, como cor e região. A abordagem 2, proposta por AGANJ et al. (2018), em que eles fazem a segmentação da imagem baseado no centro de massa local e a avalia usando esse próprio atributo da imagem.

Neste trabalho terá o foco em avaliar a segmentação se preocupando apenas com a entrada e a saída dos dados, conhecida como teste de caixa preta. Esse tipo de teste é definido como uma técnica de teste na qual a funcionalidade do aplicativo em teste é testada sem examinar a estrutura interna do código, os detalhes da implementação e o conhecimento dos caminhos internos do software. Esse tipo de teste é baseado inteiramente em requisitos e especificações de software. No BlackBox Testing, focamos apenas nas entradas e saídas do sistema de software sem nos preocuparmos com o conhecimento interno do programa (Guru99, acesso em 01/03/2020).

Dado que a maioria das bases de dados disponíveis não possuem imagens de referência para validar a segmentação, esse estudo tem o foco em abordagens não supervisionadas. Nossa proposta foi conduzir um estudo sobre métricas de validação não supervisionadas, através de suas aplicações à *datasets* de referência, cujo resultado final nos trouxe resultados de avaliação para ajudar aos desenvolvedores de algoritmo de segmentação de imagem. Como alvo da aplicação das abordagens não segmentadas, foram selecionadas as métricas de avaliação não supervisionadas para imagens coloridas, foram encontradas as métricas F, Q, Frc e E, por não ter muita documentação não foi possível implementar todas, e foram selecionados dois conjuntos de imagens (*datasets*): (1) *Berkeley Segmentation Data Set* (BSDS) (EECSBerkeley, 2007), que contém imagens naturais, representando um cenário menos crítico; (2) uma base de feridas médicas, com o intuito de avaliar um cenário mais crítico.

1.3 Objetivos

O objetivo geral deste trabalho foi avaliar métricas de validação não supervisionadas em imagens segmentadas, independentes do modelo de segmentação utilizado. Para alcançar esse objetivo, propomos os seguintes objetivos específicos:

- Realizar uma revisão bibliográfica sobre os tipos de avaliação de segmentação;
- Selecionar as métricas de avaliação não supervisionadas;
- Avaliar a aplicação das métricas selecionadas aos conjuntos de imagens citados anteriormente.

1.4 Estrutura do documento

O trabalho foi dividido em oito capítulos. No capítulo 2 foi discutido o referencial teórico para o estudo deste trabalho, discutindo-se aprendizado de máquina, segmentação e avaliação de segmentação; no capítulo 3, discute-se as métricas de avaliação supervisionada e não supervisionada; no capítulo 4, explicamos como foi aplicado as métricas de avaliação; no capítulo 5 mostramos os resultados da avaliação dos conjuntos de dados, tanto do BSDS quanto do banco de imagens de feridas; no capítulo 6 temos a conclusão do estudo; no 7 nossas referências e no capítulo 8, um apêndice.

2 REFERENCIAL TEÓRICO

Neste capítulo será discutido sobre os assuntos que levaram à criação deste trabalho. Falaremos sobre aprendizado de máquina, segmentação de imagem e tipos de avaliação de segmentação de imagem.

2.1 Aprendizado de Máquina

Segundo Coppin (2010), o aprendizado está diretamente ligado com a inteligência, pois se um sistema é, de fato, capaz de aprender a exercer determinada tarefa, merece então ser chamado de inteligente. De acordo com Santos (2005), o aprendizado de máquina nada mais é que um aprendizado por experiência, que conforme a tarefa é executada, o problema aprende a melhor maneira de resolver. Além de estruturar o conhecimento existente, que leva a um entendimento do aprendizado.

Nesse contexto, Neves (2018) fala que o aprendizado de máquina é a área da inteligência artificial responsável pelos métodos e algoritmos que possuem a capacidade de aprender com informações obtidas através de uma base de dados. O conceito desta forma de cognição difere do processo cognitivo humano, porém, se baseia na mesma. Isto parte do princípio de que através do erro se consegue melhorar uma dada resposta, ou seja, consiste em acreditar que o aprendizado é adquirido através da experiência.

Os algoritmos desta área podem ser divididos em mais de uma categoria, dentre elas podemos citar a aprendizagem de máquina supervisionado, não supervisionado e semi-supervisionado, sendo:

- Aprendizado de Máquina Supervisionado é o termo usado sempre que o programa é “treinado” sobre um conjunto de dados pré-definido. Baseado no treinamento com os dados pré-definidos, o programa pode tomar decisões precisas quando recebe novos dados (Ciências e Dados, acesso 19/07/2019).
- Aprendizado de Máquina Não Supervisionado é o termo utilizado quando um programa pode automaticamente encontrar padrões e relações em um conjunto de dados (Ciências e Dados, acesso 19/07/2019).

- Aprendizado de Máquina Semi-Supervisionado utiliza exemplos rotulados para se obter informações sobre o problema e usá-la para guiar o processo de aprendizado a partir dos exemplos não rotulados (BRUCE, 2001).

Um exemplo de uso de aprendizagem de máquina é no trabalho de SOUZA et al., 2015 onde ele usa um perceptron de múltiplas camadas para poder classificar uma arritmia cardíaca.

2.1.1 Segmentação de Imagem

Segundo Zhang et al. (1996), em um sentido convencional, a segmentação de imagem é o particionamento de uma imagem em regiões, em que as partes de dentro de uma região são semelhantes de acordo com algum predicado de uniformidade e diferentes entre regiões vizinhas. Devido à sua importância, muitos algoritmos de segmentação e critérios de avaliação foram propostos. Apesar disso, pouquíssimos estudos comparativos sobre os métodos utilizados para avaliação foram publicados. A segmentação de imagens é uma etapa essencial para muitos aplicativos avançados de criação de imagens e considerado um dos difíceis problemas de reconhecimento e entendimento de imagens (SHETA et al., 2014).

Figura 1 - Exemplos de segmentação



Fonte: TANNER, G (2019)

Na imagem acima é possível ver um exemplo de uma segmentação de imagem, onde vários elementos são segmentados como as árvores, a faixa de pedestre, os carros e as pessoas

na imagem. Os objetos mais próximos na imagem são facilmente segmentados já os que estão a uma distância maior são confundidos com outros objetos.

2.2 Avaliação de Segmentação

Normalmente, os pesquisadores mostram seus resultados de segmentação em algumas imagens e apontam que os resultados parecem bons. Todavia, nunca sabemos de tais estudos se os resultados são bons ou exemplos típicos.

Muitos métodos de segmentação de imagens foram propostos nas últimas décadas como segmentação através da maximização da entropia difusa (YIN et al., 2017) onde ele cria uma árvore binária com dois níveis e a cada nível é aplicado o Erro quadrático médio, onde o erro for menor a árvore continua descendo até chegar na melhor segmentação, segmentação com base no centro de massa local (AGANJ et al., 2018) onde é feito a segmentação da imagem e em seguida é comparado com outro método de segmentação o centro de massa local da imagem entre outros métodos. Com os novos modelos de segmentação propostos, diversos métodos de avaliação foram utilizados para comparar novos modelos de segmentação com os anteriores. Esses métodos são fundamentalmente muito diferentes e são divididos em métodos objetivos que são baseados na avaliação dos resultados da segmentação por juízes humanos, e métodos subjetivos que são divididos em analíticos, supervisionado e não-supervisionado.

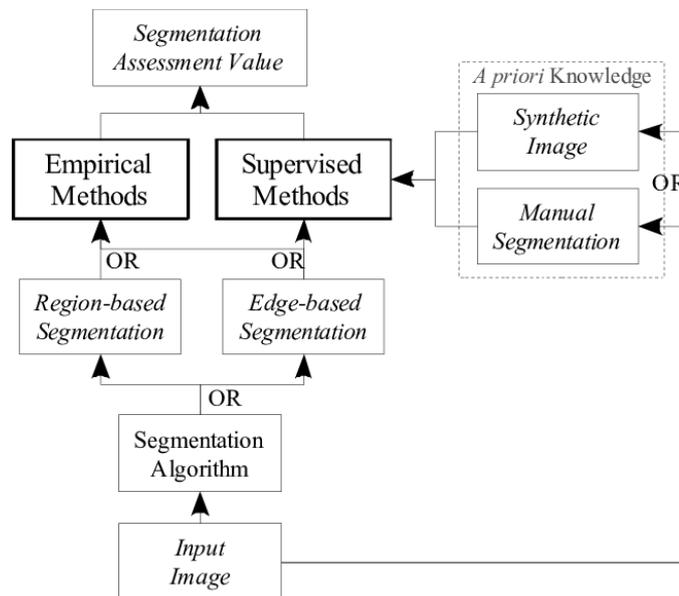
2.2.1 Método Analítico

Segundo Zhang et al. (1996), os métodos analíticos avaliam algoritmos de segmentação independentemente de sua saída, avaliando-os com base em certas propriedades dos algoritmos de segmentação, como estratégia de processamento (paralela, sequencial, iterativa ou mista), complexidade de processamento, eficiência de recursos e resolução de segmentação, que geralmente não são considerados eficazes para avaliar a sua qualidade, por exemplo em Técnicas de avaliação Limiaries (WESZKA; ROSENFELD, 1978), em que se aplica o método avaliativo para validar propriedades do algoritmo. Essas propriedades geralmente são independentes da qualidade dos resultados de segmentação de um algoritmo, portanto, essas propriedades não são consideradas eficazes para caracterizar a diferença de desempenho entre os algoritmos de segmentação.

2.2.2 Método Supervisionado

Yang et al. (1995) falam que os métodos supervisionados avaliam algoritmos de segmentação comparando a imagem segmentada resultante com uma imagem de referência segmentada manualmente. O grau de similaridade entre as imagens segmentadas humana e máquina determina a qualidade da imagem segmentada. Esses métodos podem ser vistos por exemplo em *Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm (2015)*, a avaliação feita neste artigo é usando os métodos de *Peak to Signal Noise Ratio (PSNR)* e *Root Mean Square Error (RMSE)*. Eles são usados a dois tipos de algoritmos de segmentação, onde o autor propõe uma segmentação usando *K-means Clustering Algorithm* e *Subtractive Clustering Algorithm*, essa segmentação proposta pelo autor é comparado com *K-means algorithm*. Este método pode ser compreendido melhor com o diagrama abaixo.

Figura 2 - Diagrama de avaliação de segmentação supervisionada



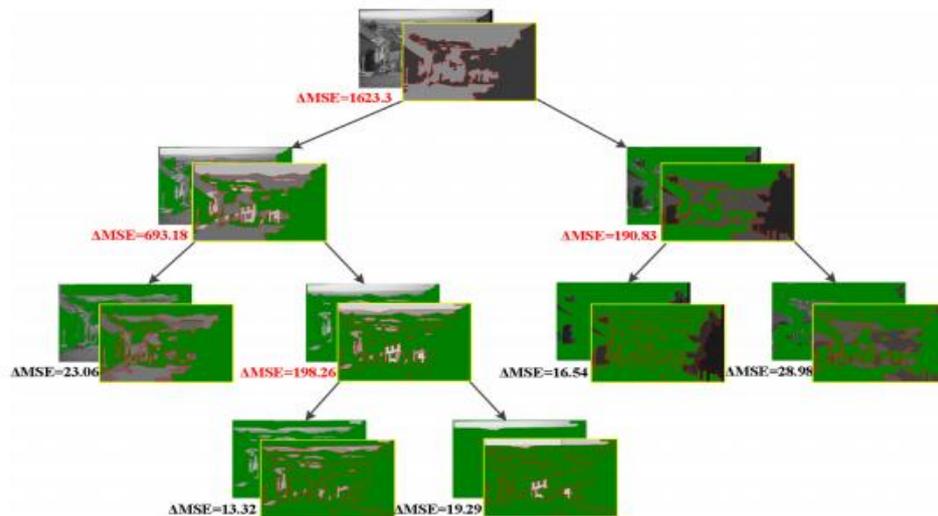
Fonte: MARTIN (2008)

A imagem de entrada é segmentada e aplicado o método de avaliação supervisionado, uma imagem de referência é aplicada junto com a imagem de entrada. No final do fluxograma é visto o valor da avaliação da segmentação da imagem de entrada em comparação com a de referência.

2.2.3 Métodos Não-Supervisionados

Enquanto os métodos supervisionados avaliam imagens segmentadas em relação a uma imagem de referência, Chabrier et al. (2004) afirmam que os métodos de avaliação não supervisionada não requerem uma imagem de referência, mas avaliam uma imagem segmentada com base em sobre quão bem ele corresponde a um amplo conjunto de características de imagens segmentadas, **conforme desejado pelos seres humanos**. Esse método pode ser observado em *Image Segmentation Evaluation: A Survey of Unsupervised Methods (2008)*, onde ele menciona várias métricas de avaliação tanto para imagens coloridas e escala de cinza. É usado as métricas para imagens coloridas em segmentações de imagens naturais e o resultado das métricas é comparado. São usados vários tipos de comparações e segmentações como segmentação humana versus segmentação da máquina, é comparado com dois modelos de segmentação diferentes e ao final ver o resultado de cada avaliação de segmentação, outro exemplo de avaliação é ilustrado na imagem abaixo.

Figura 3 - Exemplo de avaliação de segmentação não supervisionada



Fonte: YIN et al. (2017).

É usado uma segmentação em hierarquia onde a cada nível da árvore ao lado direito posiciona a imagem mais escura e a esquerda a mais clara, é usado o erro quadrático médio (MSE) para fazer a avaliação da segmentação, quanto maior seu valor melhor é o resultado se sua segmentação.

2.2.4 Métricas de Avaliação Não Supervisionadas

Cada métrica aplicado neste trabalho existe uma fórmula que é usada para sua avaliação. As formulas usadas das métricas será discutido abaixo:

- Medida **F** é calculada da seguinte forma:

$$F(I) = \sqrt{R} \times \sum_{i=1}^R \frac{e_i^2}{\sqrt{A_i}}$$

Fonte: LIU (1994)

Onde I é a imagem a ser segmentada, R é o número de regiões na imagem segmentada, A_i a área ou o número de pixels da i -ésima região e e_i o erro de cor da região i . E e_i é definida como a soma da distância euclidiana dos vetores de cores entre a imagem original e a imagem segmentada de cada pixel na região. O termo $\overline{A_i}$ é uma medida global que penaliza a segmentação que forma muitas regiões. O termo $\frac{e_i^2}{A_i}$ é uma medida local que penaliza pequenas regiões ou regiões com um grande erro de cor e_i indica se uma região está ou não atribuída a um recurso apropriado (cor).

- A medida **Q** é calculada da seguinte forma:

$$Q(I) = \frac{1}{10000(N \times M)} \sqrt{R} \times \sum_{i=1}^R \left[\frac{e_i^2}{1 + \log A_i} + \left(\frac{R(A_i)}{A_i} \right)^2 \right],$$

Fonte: BORSOTTI et al. (2001)

Onde o termo \overline{R} é uma medida global que penaliza a segmentação que forma muitas regiões, $R(A_i)$ representa o número de regiões com uma área igual a A_i , $N \times M$ é o tamanho da imagem e e_i o erro de cor da região i .

- A medida **F_{RC}** é calculada da seguinte forma:

$$F(\underline{D}(I^j), \overline{D}(I^j)) = \frac{\overline{D}(I^j) - \underline{D}(I^j)}{2}$$

Fonte: ROSENBERG (2001)

Onde \underline{D} é a medida intra-região e \overline{D} é a medida inter-região. A medida intra-região é calculada da seguinte forma:

$$\underline{D}(I^j) = \frac{1}{m^j} \sum m^j \frac{r_i^j}{NT} \underline{D}(R_i^j)$$

Fonte: ROSENBERG (2001)

Onde m^j é o número de regiões de R_i^j , r_i^j é o número de pixel na região R_i^j e NT o número total de pixel da imagem. O cálculo da medida inter-região é similar.

- A medida E é calculada da seguinte forma:

$$E = H_\ell(I) + H_r(I).$$

Fonte: ZHANG et al. (2004)

Onde $H_r(I)$ é a entropia esperada da região e $H_\ell(I)$ é a entropia de layout. A entropia esperada da região é representada da seguinte maneira:

$$H_r(I) = \sum_{j=1}^N \left(\frac{S_j}{S_I} \right) H(R_j).$$

Fonte: ZHANG et al. (2004)

Onde S_j a área ou o número de pixels da j -ésima região, S_I a área ou número de pixel da imagem e $H(R_j)$ é a entropia por região. A entropia de layout é representada da seguinte forma:

$$H_\ell(I) = - \sum_{j=1}^N \frac{S_j}{S_I} \log \frac{S_j}{S_I}.$$

Fonte: ZHANG et al. (2004)

Onde S_j a área ou o número de pixels da j -ésima região, S_l a área ou número de pixel da imagem.

O Restante das equações das métricas não supervisionadas podem ser vistas no Apêndice.

3 MÉTRICAS DE AVALIAÇÃO DE SEGMENTAÇÃO

Neste capítulo será discutido sobre as métricas de avaliação que foram estudadas, e explicado como foi possível encontrar as métricas: avaliação supervisionadas e não supervisionadas.

3.1 Revisão Bibliográfica

O estudo para chegar até as métricas de avaliação foi concluído lendo os artigos de segmentação e vendo como era feito a avaliação da segmentação. Outro parâmetro que foi observado antes de selecionar as métricas de avaliação, foi ver se o método era usado sem precisar ter acesso ao código da segmentação, se preocupando só com a saída. Esse tipo de procedimento é chamado de teste de caixa preta, que segundo Myers (1979), para esse método é preciso visualizar o programa como uma caixa preta. Seu objetivo é não se preocupar com o comportamento e a estrutura interna do programa, ou seja, se preocupar somente com a entrada de dados e a saída dos dados.

3.2 Métricas de Avaliação Supervisionadas

Vários métodos de avaliação de segmentação de imagens foram propostos ao longo dos anos. No entanto, todos eles são baseados em diferentes princípios funcionais subjacentes e fazem suposições diferentes sobre os dados / informações da imagem. Como resultado, essas medidas têm bom desempenho em alguns casos e mal em outros. Assim, cada método tem suas vantagens e limitações (ZHANG, 1996). Portanto, é dever dos desenvolvedores do sistema escolher as medidas adequadas à sua aplicação. As métricas a seguir são usadas em uma imagem segmentada a qual deseja avaliar sua segmentação e aplicada em uma imagem de referência, é feito uma comparação entre o resultado de ambas as imagens e ver se a imagem foi bem segmentada ou não. A seguir, são apresentadas algumas das medidas disponíveis na literatura:

- **H2** - Compara a semelhança / dissimilaridade entre seus recursos HOG e Harris usando a distância euclidiana (SIMFUKWE et al., 2017).
- **HOSUR** - Compara similaridade/dissimilaridade entre suas características HOG e SURF usando a distância euclidiana (SIMFUKWE et al., 2017).

- **Boudary Displacement Error (BDE)** - Mede o erro médio de deslocamento dos pixels de limite entre duas imagens segmentadas (FREIXENET et al., 2002).
- **Probability Rand Index (PRI)** - Medida de agrupamento mede a semelhança entre dois agrupamentos de dados (PANTOFARU; HEBERT, 2005).
- **Variation of Information (VOI)** - É uma medida da distância entre dois agrupamentos (MEILA, 2005).
- **Global Consistency Error (GCE)** - Pressupõe que uma das segmentações deve ser um refinamento da outra e força os refinamentos locais a seguirem a mesma direção (MARTIN et al., 2001).
- **Local Consistency Error (LCE)** - O erro de consistência local permite que refinamentos ocorram nos dois sentidos em locais diferentes na segmentação (MARTIN et al., 2001).
- **Hausdorff Distance (HD)** - Avalia segmentações comparando as formas dos segmentos em uma determinada segmentação S e seu respectivo verdade terrestre G (HUTTENLOCHER et al., 1993).
- **Dice Index (DI)** - Dada uma segmentação S e sua verdade solo G , o DI é definido para comparar as regiões correspondentes de S e G (DIETENBECK et al., 2010).
- **F1 Score** - Avalia uma segmentação S em termos do grau de sobreposição entre os seus objetivos e as da respectiva verdade solo G (site <http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/glascontest/evaluation>, acesso em 13 de fevereiro de 2020).
- **F-Measure** - É a média harmônica entre precisão e revogação. Uma vez que seu valor está alto significa que a acurácia que obtivemos é relevante (ALPERT et al., 2007).
- **PSRE** - A relação pico/ruído do sinal é a proporção entre as potências máximas atingíveis e o ruído corrompido que influência a semelhança da imagem. É usado para medir a qualidade da imagem de saída (KAUR, 2012).
- **RMSE** - Medida de precisão para comparar erros de previsão de diferentes modelos para um conjunto de dados específicos e não entre conjuntos de dados, pois depende da escala (KAUR, 2012).

3.3 Métricas de Avaliação Não-Supervisionadas

Embora os métodos supervisionados sejam os métodos de avaliação quantitativa objetiva mais amplamente utilizados, alguns métodos não supervisionados foram propostos. Muitos dos métodos anteriores nessa área focavam apenas a avaliação da segmentação de primeiro plano em segundo plano ou apenas imagens em nível de cinza. No entanto, muitos desses métodos contêm teoria que é benéfica para imagens com vários segmentos e podem ser adaptados à avaliação de segmentação de imagens em cores revisitando a teoria fundamental e reprojutando os métodos de acordo com as novas restrições (ZHANG et al., 2008). Essas métricas de avaliação são usadas na imagem de entrada e na de saída, de acordo com o resultado da avaliação vai dizer se a imagem foi bem segmentada ou não. A seguir algumas métricas de avaliação não supervisionadas:

- **D_{WR}** - Mede a diferença de nível de cinza entre a imagem original e a imagem de saída após o limiar (WESZKA; ROSENFELD, 1978).
- **Busy** - Baseia-se na medida de "ocupação" na imagem, com a suposição de que os objetos e o plano de fundo ideais não são fortemente texturizados e têm formas compactas simples (WESZKA; ROSENFELD, 1978).
- **η** - Mede a variação intra e inter-região do objeto em primeiro plano e o plano de fundo (OTSU, 1979).
- **PV** - É um conjunto de medidas de segmentação que constituem um vetor de desempenho (PV). O vetor PV armazena os fatores que caracterizam a segmentação, incluindo uniformidade da região, contraste da região, contraste da linha, conectividade da linha e textura (LEVINE; NAZIF, 1985).
- **NU** - Medida que aprimora a medida de uniformidade da região na PV para usar uma medida de uniformidade da região normalizada (SAHOO et al., 1988).
- **SM** - É uma medida de forma. É definida como a soma dos gradientes em cada pixel cujo valor do recurso excede o limite de segmentação e o valor médio de seus vizinhos (SAHOO et al., 1988).
- **SE** - É uma medida de avaliação de segmentação baseada em entropia para uniformidade intra-região com base na entropia local de segunda ordem (PAL; BHANDARI, 1993).

- **F** - É um critério de avaliação que mede o erro médio de cor ao quadrado dos segmentos, penalizando a segmentação excessiva ponderando proporcionalmente à raiz quadrada do número de segmentos (LIU; YANG, 1994).
- **F'** - Foi proposto para melhorar F, porque descobriu-se que F tinha um viés em relação à segmentação excessiva, que é a característica de produzir muito mais regiões do que o desejado em um único objeto do mundo real (BORSOTTI et al., 1998).
- **Q** - Melhora o F e F' diminuindo o viés em relação à segmentação excessiva e subsegmentada (BORSOTTI et al., 1998).
- **F_{RC}** - É um critério de avaliação que leva em consideração a disparidade global intra-região que quantifica a homogeneidade de cada região na imagem resultante. Da mesma forma, a disparidade global inter-regiões mede a disparidade entre regiões (ROSENBERGER; CHEHDI, 2000).
- **Z_{eb}** - É um critério de avaliação baseado no contraste interno e externo das regiões medidas na vizinhança de cada pixel (CHABRIER et al., 2004).
- **E_{cw}** - É um método de avaliação composto para imagens coloridas. Ele usa o erro visual intra-região para avaliar o grau de sub-segmentação e usa o erro visual da região inter-região para avaliar o grau de super-segmentação (CHEN; WANG, 2004).
- **E** - É uma função de avaliação baseada no tamanho mínimo da descrição (MDL - minimum description length). Ela usa a entropia de região e a de layout. Na primeira, ele usa a entropia da região como sua medida de uniformidade intra-região, que mede a entropia das intensidades de pixel dentro de cada região. Na segunda, ele usa a entropia indicando quais pixels pertencem a quais regiões, para penalizar a segmentação excessiva quando a região entropia torna-se pequeno (ZHANG et al., 2004).
- **Mean Squared Error (MSE)** - Mede a média dos quadrados dos erros, ou seja, a diferença quadrática média entre a estimativa de valores e o valor real (GONG; YANG, 2001).

No Quadro 1 abaixo podemos ver um sumário de todas as métricas tanto supervisionada quanto não supervisionadas.

Quadro 1 – Sumário das Métricas

Métricas Supervisionadas	H2
	HOSUR
	BDE
	PRI
	GCE
	LCE
	VOI
	HD
	DI
	F1-Score
	F-Measure
	MAE
	PSRE
	RMSE
Métricas Não-Supervisionadas	<i>D_{WR}</i>
	Busy
	η
	PV
	NU
	SM
	SE
	F
	F'
	Q
	<i>F_{RC}</i>
	<i>Z_{eb}</i>
	<i>E_{cw}</i>
	E
	MSE

Fonte: Elaborado pelo autor, 2020.

Neste capítulo foram discutidos como foi possível chegar até todas as métricas de avaliação, nas métricas supervisionadas foi visto que elas tem um bom desempenho em alguns casos e cada métrica tem sua vantagem e desvantagem, vai depender do desenvolvedor escolher qual é a métrica mais adequada ao caso. Já nas métricas não supervisionadas foi visto que muitos métodos criados avaliam em primeiro plano a imagem ou em segundo plano ou apenas em escala de cinza, contudo muitos desses métodos contêm teoria que é benéfica para imagens com vários segmentos e pode ser adaptada a imagens de cores. Além disto foi visto como é aplicado as métricas de avaliação em casa um dos métodos (Supervisionado e Não Supervisionado).

4 APLICANDO MÉTRICAS DE AVALIAÇÃO NÃO SUPERVISIONADAS

Neste capítulo será discutido como foram aplicadas as métricas não supervisionadas às imagens segmentadas e por que o uso deste tipo de avaliação.

No início do estudo foi primeiro usada a avaliação supervisionada, mas com os resultados das métricas não tivemos uma imagem de referência que possamos comparar e dizer se a segmentação foi boa ou não. Visto isso, passamos a usar os métodos de avaliação não-supervisionada para nosso estudo.

As imagens usadas para o estudo foram do tipo colorida então foi preciso selecionar métricas de avaliação não supervisionada voltadas para imagens coloridas. Para a avaliação da segmentação, nas imagens segmentadas foram usadas as métricas de avaliação F, Q, F_{RC} e E, pois, foram as únicas que métricas que foram encontradas durante a pesquisa. Não foi possível implementar todas as outras métricas por falta de documentação.

Para ser feita a avaliação foi usado um código feito em C++ que foi criado por Lu et al. (2016). Foi usado o Visual Studio 2013 junto com o OpenCV versão 2.1.9, e o código foi executado para cada uma das imagens.

5 ESTUDOS DE CASO

5.1 Berkeley Segmentation Data Set

É um banco de dados da *University California Berkeley* que tem como objetivo fornecer uma base empírica para pesquisas sobre segmentação de imagens e detecção de limites. O conjunto de dados usado neste banco de dados são imagens naturais que foram segmentadas manualmente. As anotações humanas servem como base para a aprendizagem de dicas de agrupamento, bem como uma referência para comparar diferentes algoritmos de segmentação e detecção de limites.

5.1.1 Resultados

A tabela 1 abaixo mostra diferentes resultados de avaliação, para a métrica F e Q, que quanto mais próximo de 0, é considerado uma boa segmentação. Já para a métrica Frc, quanto maior o seu resultado, melhor a segmentação e para a métrica E diferentemente de F e Q, não é minimizada quando a imagem é segmentada ao máximo com um pixel por região. Nesse caso, a entropia do layout se torna muito grande, ou seja, quanto menor o resultado de E, significa que é uma boa segmentação.

Tabela 1 - Resultados do BSDS

BSDS500				
	F	Q	Frc	E
1	1,38075	0,369605	38,5707	2,38951
2	1,31381	0,445249	27,2634	3,2769
3	1,35174	0,569671	25,1165	3,7361
4	1,59022	0,546323	24,6284	3,63098
5	0,932042	0,420481	31,6927	3,77695
6	1,77766	0,673446	28,755	3,07316
7	1,85651	0,788374	22,8787	3,74978
8	1,98933	0,82782	16,3278	3,96417

9	1,37438	0,535876	28,35	3,75443
10	0,536026	0,317502	21,3162	3,22798
11	3,825	1,57596	16,4797	3,81147
12	0,673408	0,273925	28,7084	3,36836
13	1,17759	0,555967	20,5992	3,91729
14	0,910044	0,455189	22,572	3,35512
15	1,37968	0,607734	19,9563	3,76186
16	1,17889	0,514197	15,8472	3,72339
17	0,862513	0,364645	17,8414	3,06921
18	1,77881	0,785984	22,3461	3,66364
19	1,31205	0,586652	14,5172	3,7554
20	0,865069	0,389692	22,1739	3,57337
21	1,5771	0,683951	25,8921	3,8036
22	1,14563	0,439091	22,9882	3,55236
23	1,02478	0,409115	22,5684	3,45096
24	0,707804	0,319527	25,1808	3,63693
25	1,69245	0,734773	16,4314	3,77594
26	0,853326	0,441244	19,4223	3,6515
27	0,834174	0,367079	14,6071	3,29246
28	0,83425	0,411286	38,6077	3,52896
29	1,5529	0,515004	34,7948	3,38772
30	0,576663	0,273975	29,7534	3,35707

Fonte: Elaborado pelo autor (2020)

5.1.2 Discussão (*Super Segmentação, Sub-Segmentação, Boa Segmentação*)

Para indicar uma boa segmentação não é preciso que todas as métricas em uma imagem tenham saído com um bom resultado, cada uma delas indica se ela foi bem segmentada de acordo com a sua descrição.

Como por exemplo a imagem abaixo, Figura 1, mostra um resultado bom para as métricas de Erro de cor médio (F e Q) e um resultado bom para a métrica de equilíbrio de entropia (E), mas para a métrica de disparidade intra-região e inter-região teve um resultado baixo.

Figura 4 - Imagem 17, ver Tabela 1 para comparar dados das métricas.



F: 0,862513 - Q: 0364645 - Frc: 17,8414 - E: 3,06921

Fonte: EECSBerkeley, (2007)

Para indicar uma segmentação excessiva ou sub-segmentação temos que olhar os valores do resultado de F e Q, quanto mais se distanciar de zero indica que ela teve uma segmentação excessiva ou foi sub-segmentada. Como por exemplo a imagem abaixo, Figura 2.

Figura 5 - Imagem 11, ver Tabela 1 para análise das métricas.



F: 3,825 - Q: 1,57596 - Frc: 16,4797 - E: 3,81147

Fonte: EECSBerkeley, (2007)

Pode se ver que no contorno das árvores na imagem segmentada (Figura 2), ele seleciona um pouco do céu. Nesta imagem temos um exemplo de segmentação excessiva. Um exemplo de uma imagem sub segmentada é visualizado a seguir, onde faltou segmentar rochas no meio da imagem para cima com a mesma coloração das outras rochas que foram contornadas (Figura 3).

Figura 6 - Imagem 7, ver Tabela 1 para análise das métricas.



F: 1,85651 - Q: 0,788374 - Frc: 22,8787 - E: 3,74978

Fonte: EECSBerkeley, (2007)

5.2 Banco de Imagens de Feridas

É um banco de dados disponível da Cicatriza, que é um centro de tratamento de feridas, onde se trata todo tipo de ferida, lesões ou úlceras. O conjunto de dados usa imagens que contêm feridas em alguns pacientes.

Para avaliação deste conjunto de dados foi usado o mesmo código mencionado na sessão anterior, executando-se o código para cada imagem e o resultado é mostrado na Tabela 2.

5.2.1 Resultados

A explicação dos resultados da tabela 2 pode ser vista na sessão de resultados do BSDS na sessão 5.1.1. Os resultados negativos serão explicados na sessão seguinte.

Tabela 2 - Resultado do Banco de Imagens de Feridas

	CICATRIZA			
	F	Q	Frc	E
Granulado 1	0,0481779	1,88641	23,7257	3,13601
Granulado 2	0,04718	-0,593705	20,28	3,20082
Granulado 3	0,0710361	2,20596	20,4237	2,83928
Granulado 4	0,0563386	0,262502	41,1773	2,96749
Granulado 5	0,0405665	-0,488393	30,9039	3,14399
Granulado 6	0,070903	-0,588741	54,3307	3,01823
Granulado 7	0,120778	-1,71023	14,3838	3,17909
Granulado 8	0,0769257	0,840299	41,3525	3,04452
Granulado 9	0,17426	-1,14316	15,2207	2,84937
Granulado 10	0,053172	1,80064	30,4904	2,96199
Granulado 11	0,0484611	-0,974746	25,9009	3,04272
Granulado 12	0,0165443	-0,0771693	16,27	2,86677
Granulado 13	0,0640411	0,22897	25,3587	3,10972

Granulado 14	0,0847921	0,979601	35,4797	3,29727
Granulado 15	0,120918	-1,81378	22,8691	3,15687
Granulado 16	0,0649172	-0,751281	12,4601	3,20217
Granulado 17	0,0487887	-1,61717	7,50462	3,06429
Granulado 18	0,0826719	3,7319	13,6026	3,11162
Granulado 19	0,061964	2,00486	16,707	3,17093
Granulado 20	0,0351443	0,381815	12,378	3,10402
Granulado 21	0,025651	0,741984	13,419	3,11472
Granulado 22	0,153635	4,44986	8,62008	3,48277
Granulado 23	0,160389	-20,8267	18,6942	3,46002
Granulado 24	0,233276	-3,8113	31,9003	3,61841
Granulado 25	0,156248	1,93045	19,334	3,55938
Granulado 26	0,108701	-31,2854	22,9407	2,99346
Granulado 27	0,0689417	1,33995	20,4204	3,0536
Granulado 28	0,0439564	-0,899844	28,7715	3,3515
Granulado 29	0,0963642	-2,2324	15,6003	3,31895
Granulado 30	0,02568	1,01249	20,3697	3,12651

Fonte: Elaborado pelo autor (2020).

5.2.2 *Discussão (Super Segmentação, Sub-Segmentação, Boa Segmentação)*

Conforme discutido na sessão de discussão dos resultados do conjunto de dados do BSDS sobre uma boa segmentação, a mesma argumentação se aplica a esses resultados. Para uma segmentação excessiva ou sub-segmentação, podemos observar apenas a métrica de avaliação Q, que quanto mais longe fica de zero, mais ela pode ser super ou sub-segmentada. Dado como exemplo de uma imagem sub-segmentada onde a segmentação só pintou a área mais clara da ferida e deixou a área mais escura sem ser segmentada (Figuras 4 e 5).

Figura 7 - Granulado 27. Ver dados das métricas na Tabela 2.



F: 0,0689417 - Q: 1,33995 - Frc: 20,4204 - E: 3,0536

Fonte: Cicatriza.

Figura 8 - Granulado 15. Ver dados das métricas na Tabela 2.



F: 0,120918 - Q: -1,81378 - Frc: 22,8691 - E: 3,15687

Fonte: Cicatriza.

Para algumas feridas que foram segmentadas e o resultados se mostra longe de zero, verificamos uma coloração um pouco parecida com a pele, observando-se que a segmentação foi excessiva, conforme na figura 5.

6 CONCLUSÃO

Neste trabalho, pesquisamos e buscamos os métodos de avaliação com objetivo de validar segmentações de imagens. Primeiramente foi preciso fazer uma pesquisa buscando meios de validar este processo. Foi preciso identificar os atributos que precisaríamos para poder validar com cada método. No meio de avaliação analítica não foi possível usar o método por não mede a qualidade da imagem, o método supervisionado precisa de uma imagem de referência para poder avaliar a segmentação. Assim, foi usado o método não supervisionado por não precisar de nenhuma referência para validar a segmentação, e neste caso só precisaríamos da imagem original e da imagem segmentada para o processo.

Os métodos de avaliação de segmentação não supervisionados oferecem a vantagem exclusiva de serem puramente objetivos e não exigirem uma imagem de referência segmentada manualmente. Essa vantagem é crucial para aplicativos de segmentação de uso geral, como aqueles incorporados em sistemas em tempo real, onde uma grande variedade de imagens com conteúdo desconhecido e nenhuma verdade básica precisam ser segmentadas.

Ao aplicar os métodos de avaliação não supervisionado F, Q, Frc e E aos conjuntos de imagens foi possível observar que algumas imagens possuem uma super segmentação ou sub-segmentação. Foi visto que as imagens que possuam uma super segmentação é por que a pele do paciente está com uma coloração semelhante ao da ferida, fazendo com que o algoritmo segmente mais do que o desejado. Já com a sub-segmentação foi visto que as imagens que têm esse parâmetro são por causa de falta de iluminação na imagem, uma sombra por exemplo. Assim, ao segmentar a imagem é visto que só seleciona a parte mais iluminada.

É preciso ajustar o algoritmo de segmentação para que ele não faça uma sub-segmentação por falta de iluminação na imagem e faça uma super segmentação por causa da coloração da pele parecida com a da ferida.

Como com este algoritmo nos preocupamos só com a entrada e a saída dos dados, para um próximo trabalho, sugere-se estudar métodos que avaliam o algoritmo da segmentação e que vejam a eficiência da sua segmentação.

REFERÊNCIAS

AGANJ, I; HARISINGHANI, M.G; WEISSLEDER, R; FISCHL, B. Unsupervised Medical Image Segmentation Based on the Local Center of Mass. **Scientific Reports**, 2018.

ALPERT. S; GALUN, M; BASRI, R; BRANDT, A. Image segmentation by probabilistic bottom-up aggregation and cue integration, in **Proc. IEEE Conf. Comput. Vis. Pattern Recognit.**, 2007, pp. 1–8.

BORSOTTI, M; CAMPADELLI, P; SCHETTINI, R. Quantitative evaluation of color image segmentation results, **Pattern Recognition Letters**, vol. 19, no. 8, pp. 741C747, June 1998.

BRUCE, R. A Bayesian Approach To Semi-Supervised Learning. 2001.

CHABRIER, S; EMILE, B; LAURENT, H; ROSENBERGER, C; MARCHE, P. Unsupervised evaluation of image segmentation application to multispectral images, in **Proc. of the 17th international conference on pattern recognition**, 2004.

CHEN, H.-C; WANG, S.-J. The use of visible color difference in the quantitative evaluation of color image segmentation, in **Proc. ICASSP**, 2004.

Cicatriz, Cicatriza | Tratamento de Feridas e Curativos, Disponível em: <https://cicatrizacg.com.br/>, Acesso em: 1 de Março de 2020

COPPIN, B. Inteligência artificial. Rio de Janeiro: LTC, 2010.

DHANACHANDRA, N; MANGLEM, K; CHANU, Y.J. Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm. **National Institute of Technology**, India, 2015.

DIETENBECK. T; ALESSANDRINI, M; FRIBOULET, D; BERNARD, O. CREASEG: a free software for the evaluation of image segmentations algorithms based on level-sets, in **Proceeding of IEEE 17th International Conference on Image Processing**, Hong Kong, China 2010, pp. 665–668.

EECSBerkeley, The Berkeley Segmentation Dataset and Benchmark. Disponível em: <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>. 2007. Acesso em: 4 de fevereiro de 2020.

FREIXENET, J; MUNOZ, X; RABA, D; MARTÍ, J; CUFI, X. Yet another survey on image segmentation region and boundary information integration, in **Proceeding of European Conference on Computer Vision**, Copenhagen, Denmark, 2002, pp. 408–422.

GONG, M; QIAN, Y; CHENG, L. Integrated foreground segmentation and boundary matting for live videos, **IEEE Trans. Image Process.** 24 (4) (2015) 1356–1370.

GONG, M; YANG, Y.-H. Genetic-based multiresolution color image segmentation, 2001, pp 141-148

Guru99, What is BLACK Box Testing? Techniques, Examples & Type. <https://www.guru99.com/black-box-testing.html>, Acesso em: 01 de Março de 2020.

HUTTENLOCHER, D.P; KLANDERMAN, G.A; RUCKLIDG, W.J. Comparing images using the Hausdorff distance, **IEEE Trans Pattern Anal Mach Intell.** 15(9) (1993), 850–863.

LEVINE, M.D; NAZIF, A.M. Dynamic measurement of computer generated image segmentations, **IEEE Transactions on Pattern Analysis and Machine Intelligence** 7 (2) (1985) 155–164.

LIU, J; YANG, Y.-H. Multi-resolution color image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 7, pp. 689C700, July 1994.

MARCO'S BLOG. Under vs OverSegmentation, Disponível em <https://marcoza.blogspot.com/2007/09/under-vs-oversegmentation.html>. Acesso em: 29 fer. 2020.

MARTIN, D; FOWLKES, C; TAL, D; MALIK, J. A database of human segmented natural images and its applications to evaluating segmentation algorithms and measuring ecological

statistics, in **Proceeding of IEEE International Conference on Computer Vision**, Vancouver, British Columbia, Canada, 2001, pp. 416–423.

MARTIN, V; THONNAT, M. **A Cognitive Vision Approach to Image Segmentation**, Tools in Artificial Intelligence, 2008.

MATOS, D. **Conceitos Fundamentais de Machine Learning**, Disponível em: <http://www.cienciaedados.com>. Acesso em: 19 jul. 2019.

MEILA, M; Comparing clusterings—an axiomatic view, in **22nd International Conference on Machine Learning**, Bonn, Germany, 2005, pp. 577–584.

MYERS, G.F. **The Art of Software Testing**. 1a edição, John Wiley and Sons, 1979.

NEVES, S. A. **Técnicas de Aprendizado de Máquina Aplicadas a Classificação da Qualidade de Pavimentos Asfálticos utilizando Smartphones**. 2018.

OTSU, N; A threshold selection method from gray-level histograms, **IEEE Transactions on Systems, Man and Cybernetics** 9 (1) (1979) 62–66.

PAL, N; BHANDARI, D. Image thresholding: some new techniques, **Signal Processing** 33 (2) (1993) 139–158.

PANTOFARU, C; HEBERT, M. A Comparison of Image Segmentation Algorithms, CMU-RI-TR-05-40, **Carnegie Mellon University**, Pittsburgh, PA, 2005, pp. 1–32.

ROSENBERGER, C; CHEHDI, K. Genetic fusion: Application to multicomponents image segmentation, in **Proc. ICASSP-4 Istanbul**, Turkey, 2000.

SAHOO, P; SOLTANI, S; WONG, A; CHEN, Y. A survey of thresholding techniques, **Computer Vision, Graphics, and Image Processing** 41 (2) (1988) 233–260.

SANTOS, C. N. dos. **Aprendizado de máquina na identificação de sintagmas nominais: o caso do português brasileiro**. Rio de Janeiro, 2005.

SIMFUKWE, M; PENG, B; LI, T. H2: fusion of HOG and Harris features for image segmentation evaluation, in **International Conference on Intelligent Systems and Knowledge Engineering (ISKE)**, 2017.

SIMFUKWE, M; PENG, B; LI, T; KUNDA, D. HOSUR: a novel measure for evaluation of image segmentation quality, in **IEEE International Conference in Information and Communication Technologies**, 2017.

TANNER, G. FastAI Image Segmentation. Disponível em: <https://towardsdatascience.com/fastai-image-segmentation-eacad8543f6f>, 2019. Acesso em: 02/03/2020.

TECUCI, G. Artificial Intelligence, 2012. Disponível em: <http://onlinelibrary.wiley.com/doi/full/10.1002/wics.200>. Acesso em: 15 de fevereiro de 2020.

Warwick, Department of Computer Science. <http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/glascontest/evaluation/>.

WESZLA, J; ROSENFELD, A. Threshold evaluation techniques, **IEEE Transactions on Systems, Man and Cybernetics** 8 (8) (1978) 622–629.

YANG, L; ALBREGTSEN, F; LONNESTAD, T; GROTTUM, P. A supervised approach to the evaluation of image segmentation methods, in **Computer Analysis of Images and Patterns**, 1995, pp. 759C765.

YIN, S; QIAN, Y; GONG, M. Unsupervised hierarchical image segmentation through fuzzy entropy maximization, 2017.

ZHANG, H; FRITTS, J. E; GOLDMAN, S. A; Image Segmentation Evaluation: A Survey of Unsupervised Methods, 2008

ZHANG, H; FRITTS, J; GOLDMAN, S. An entropy-based objective evaluation method for image segmentation, in **Proc. SPIE- Storage and Retrieval Methods and Applications for Multimedia**, 2004

ZHANG, Y.J. A survey on evaluation methods for image segmentation, **Pattern Recognition**, 29(8):1335-1346, (1996).

APÊNDICE - Equações das métricas de avaliação Não Supervisionadas

- D_{WR} :

$$Discrepancy = \sum_i^{I_h} \sum_j^{I_w} (C_{gl}(i, j) - L(i, j))$$

Onde $C_{gl}(i, j)$ é o valor no nível de cinza do pixel $p(i, j)$ na imagem original e $L(i, j)$ é o valor do nível de cinza do pixel $p(i, j)$ na imagem após a limiar.

- E_{CW}

$$E_{intra} = \frac{\sum_{p \in I} \mu (\|C_x^o(p) - C_x^s(p)\|_{L^*a*b} - TH)}{S_I}$$

Onde $C_x^o(p)$ e $C_x^s(p)$ são valor do recurso de pixel (componentes de cor no espaço CIE L * a * b) para o pixel p na imagem original e segmentada, respectivamente S_I é a área da imagem completa. TH é o limiar para julgar diferença significativa e $\mu t = 1$ quando $t > 0$ de outra forma $\mu t = 0$.

$$E_{inter} = \sum_{i=1}^N \sum_{j=1, j \neq i}^N [\mu (TH - \|C_x^o(p) - C_x^s(p)\|_{L^*a*b}) \cdot w_{ij} / (S_I \cdot Z)]$$

Onde S_I é a área da imagem completa, Z é o fator de normalização, TH é o limiar para julgar diferença significativa. Além disso, como na equação E_{intra} , $C_x^o(p)$ e $C_x^s(p)$ são valor do recurso de pixel (componentes de cor no espaço CIE L * a * b) para o pixel p na imagem original e segmentada. $\mu t = 1$ quando $t > 0$ de outra forma $\mu t = 0$.

$$E_{CW} = E_{intra} + E_{inter}$$

- Z_{eb}

$$I_j = \frac{1}{S_j} \sum_{s \in R_j} \max\{\text{contrast}(s, t), t \in W(s) \cap R_j\}$$

Onde $W(s)$ é o vizinho de s e $\text{contrast}(s, t) = |C_x(s) - C_x(t)|$ é o contraste do pixel s e t . S_j é a área da região j e R_j é o conjunto de pixel na região j .

$$E_j = \frac{1}{N_b(R_j)} \cdot \sum_{s \in n(R_j)} \max\{\text{contrast}(s, t), t \in W(s), t \notin R_j\}$$

Onde $n R_j$ é o conjunto de pixels na borda de R_j e $N_b(R_j)$ é o tamanho total da borda de R_j . O $\text{contrast } s, t = |C_x s - C_x t|$ é o contraste do pixel s e t .

$$Z_{eb} = I_j/E_j$$

- η

$$\sigma_W^2 = \frac{S_b}{S_I} e_{gl}^2(R_b) + \frac{S_o}{S_I} e_{gl}^2(R_o).$$

Onde o subscrito "gl" indica nível de cinza, o subscrito "o" significa medidas para o objeto em primeiro plano e o subscrito "b" significa para segundo plano. S_b e S_o são as áreas dos objetos em primeiro e segundo plano, S_I é a área da imagem completa e $e_{gl}^2(R_b)$ é o erro quadrado do nível de cinza de cada região.

$$\sigma_B^2 = \frac{S_b}{S_I} \cdot \frac{S_o}{S_I} \cdot (\hat{C}_{gl}(R_o) - \hat{C}_{gl}(R_b))^2$$

Onde S_b e S_o são as áreas dos objetos em primeiro e segundo plano, S_I é a área da imagem completa, $C_{gl}(R_o)$ e $C_{gl}(R_b)$ são o valor médio do componente gl na região b e o .

$$\eta = \sigma_W^2/\sigma_B^2$$

- NU

$$NU = 1 - \frac{e_{gl}^2(R_o) + e_{gl}^2(R_b)}{Z}$$

Onde o subscrito "gl" indica nível de cinza, o subscrito "o" significa medidas para o objeto em primeiro plano e o subscrito "b" significa para segundo plano, $e_{gl}^2(R_b)$ e $e_{gl}^2(R_o)$ são os erros quadrado dos níveis de cinza de cada região e Z é o fator de normalização.

- **SE**

$$H^{(2)}(R_k) = - \sum_{i=0}^T \sum_{j=0}^T p_{ij} \ln p_{ij}$$

Onde p_{ij} é a probabilidade da matriz de coocorrência para intensidades de pixel i e j , e T é o limiar assumido.

- **SM**

$$SM = \frac{1}{C} \sum_{(x,y)} \{Sgn[C(x,y) - C_{N(x,y)}] \delta(x,y) \cdot Sgn[C(x,y) - T]\}$$

Onde $\delta(x,y)$ é o gradiente do pixel x,y , T é o limiar da segmentação e $C_{N(x,y)}$ é o valor médio dos vizinhos de pixel x,y .

- **PV** (Medindo uniformidade do nível de cinza, Medindo textura e Contraste da região)
 - Medindo uniformidade do nível de cinza

$$U = 1 - \sum_{j=1}^N \frac{e_{gl}^2(R_j) \times W_j}{Z}$$

Onde o subscrito "gl" indica nível de cinza, $e_{gl}^2(R_j)$ são os erros quadrado dos níveis de cinza de cada região, Z é o fator de normalização e W_j é um fator de ponderação.

- Medindo Textura

$$R_\alpha = \frac{NR_\alpha / S_\alpha}{NR_I / S_I}$$

Onde NR_α significa o número de regiões na área α , S_I é a área da imagem completa e S_α é a área completa de α .

- Contraste da Região

$$C_{\alpha} = \left(\sum_{R_j \in \alpha} v_j \sum_{adj R_i} p_{ij} \frac{|\hat{C}(R_i) - \hat{C}(R_j)|}{\hat{C}(R_i) + \hat{C}(R_j)} \right) / \sum_{R_j \in \alpha} v_j$$

Onde p_{ij} é o valor de adjacência usado para ponderar o contraste entre regiões, v_j é o peso da região R_j , usando uma função que aproxima a curva sensível ao contraste humano, $C(R_i)$ e $C(R_j)$ são o valor médio na região i e j .