



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS DE SOBRAL**  
**DEPARTAMENTO DE COMPUTAÇÃO**  
**CURSO DE GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO**

**MANOEL VILELA MACHADO NETO**

**SEGMENTAÇÃO SEMI-SUPERVISIONADA DE IMAGENS ATRAVÉS DE  
DINÂMICAS COLETIVAS EM REDES COMPLEXAS**

**SOBRAL**

**2023**

MANOEL VILELA MACHADO NETO

SEGMENTAÇÃO SEMI-SUPERVISIONADA DE IMAGENS ATRAVÉS DE  
DINÂMICAS COLETIVAS EM REDES COMPLEXAS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia da Computação do Campus de Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia da Computação.

Orientador: Prof. Dr. Jarbas Joaci de Mesquita Sá Junior

SOBRAL

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

M1s MACHADO NETO, MANOEL VILELA.  
Segmentação Semi-Supervisionada de Imagens através de Dinâmicas Coletivas em Redes Complexas /  
MANOEL VILELA MACHADO NETO. – 2023.  
61 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Sobral,  
Curso de Engenharia da Computação, Sobral, 2023.  
Orientação: Prof. Dr. Jarbas Joaci de Mesquita Sá Junior.

1. segmentação interativa. 2. aprendizagem semi-supervisionada. 3. redes complexas. 4. dinâmicas  
coletivas. 5. superpixels. I. Título.

CDD 621.39

---

MANOEL VILELA MACHADO NETO

SEGMENTAÇÃO SEMI-SUPERVISIONADA DE IMAGENS ATRAVÉS DE  
DINÂMICAS COLETIVAS EM REDES COMPLEXAS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia da Computação do Campus de Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia da Computação.

Aprovada em: 12 de Dezembro de 2023

BANCA EXAMINADORA

---

Prof. Dr. Jarbas Joaci de Mesquita Sá  
Junior (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Iális Cavalcante de Paula Júnior  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Wendley Souza da Silva  
Universidade Federal do Ceará (UFC)

Dedico esta monografia à minha mãe, que sempre me apoiou na minha educação e especialmente nos primeiros anos em Sobral, os quais foram tão difíceis, à minha irmã, que me auxiliou financeiramente durante o curso pré-vestibular e à minha esposa Taiene Francêz, que sempre esteve ao meu lado.

## AGRADECIMENTOS

Ao meu orientador Prof. Dr. Jarbas Joaci Mesquita de Sá Junior (UFC), pela inspiração que foi pra mim, às suas inesquecíveis aulas na UFC, além da confiança que depositou em mim e, especialmente, agradeço as inúmeras oportunidades que me deu para recomeçar quando minha vida pessoal se tornou um grande obstáculo.

Ao Prof. Dr. Filipe Verri (ITA) por ter contribuído fundamentalmente na elaboração da linha de pesquisa onde este trabalho se encaixa, pelas discussões muito inspiradoras e por suas sugestões que foram extremamente úteis para o desenvolvimento da técnica apresentada neste trabalho.

Aos meus professores da UFPA e UFC, que me compartilharam seus conhecimentos e permitiram eu construir uma base de conhecimento sólida, a qual foi muito importante para a confecção deste trabalho.

Aos meus amigos, não citarei nomes, mas que se mostraram presentes na minha vida durante a graduação, e não apenas os que fiz na UFC, mas incluo aqueles que cruzaram fronteiras. Para esses amigos, agradeço especialmente pelos momentos em que estivemos conversando por anos, ao trocar as ideias mais malucas sobre computação e os desafios nas nossas vidas pessoais, mas também por terem participado dessa jornada junto comigo e por me fazer lembrar de que eu não estava sozinho nos momentos mais difíceis.

À minha família, especialmente minha mãe e minha irmã, por ter me incentivado e apoiado na busca de uma educação de ensino superior, mesmo com todas as dificuldades que precisei enfrentar.

Em especial, agradeço a minha esposa Taiene Francêz, por ter acreditado que eu fosse capaz de superar todas os obstáculos enfrentados na reta final da graduação, incluindo os mais difíceis que envolveram a minha saúde mental.

Não menos importante, agradeço meu gato Antares pelas mordidas inspiradoras e também pela companhia durante as madrugadas no processo de escrita desta monografia.

Por fim, agradeço à empresa Neoway por ter me oferecido suporte financeiro e moral durante quase toda a graduação na UFC.

“Quanto mais aumenta nosso conhecimento,  
mais evidente fica nossa ignorância.”

(John F. Kennedy)

## RESUMO

A segmentação de imagem é uma técnica que divide a imagem em regiões de interesse, tais como objetos em uma paisagem. Os algoritmos de segmentação de imagem apresentam variações em seus tipos de aprendizado, incluindo não supervisionado, supervisionado e semi-supervisionado. No contexto de segmentação interativa, o desafio é segmentar objetos do plano de fundo com a ajuda de rótulos iniciais fornecidos por um usuário. Os superpixels são algoritmos de segmentação não supervisionada utilizados como pré-segmentação para diversos problemas de imagem, como classificação e segmentação. As redes complexas são grafos com estruturas não triviais usadas para representar certos domínios de dados, como regiões de uma imagem e suas vizinhanças. A dinâmica coletiva em uma rede complexa refere-se ao comportamento emergente e interativo de vários elementos ou atores dentro de uma rede interconectada e complexa, onde as ações de um elemento podem influenciar as ações dos outros. Neste trabalho, propomos um algoritmo de segmentação de imagem semi-supervisionado que combina as técnicas de superpixels, redes complexas e dinâmicas coletivas. O método foi avaliado em várias condições usando a base de dados GrabCut para segmentação interativa. Nosso método se mostrou tão bom quanto os melhores métodos de segmentação interativa baseados em grafos e na métrica precisão superou todos os métodos comparados.

**Palavras-chave:** redes complexas; segmentação de imagens; aprendizado semi-supervisionado; superpixel; dinâmicas coletivas; segmentação interativa.

## ABSTRACT

Image segmentation is a technique that divides the image into regions of interest, such as objects in a landscape. Image segmentation algorithms present variations in their types of learning, including unsupervised, supervised, and semi-supervised. In the context of interactive segmentation, the challenge is to segment objects from the background with the help of initial labels provided by a user. Superpixels are unsupervised segmentation algorithms used as pre-segmentation for various image problems, such as classification and segmentation. Complex networks are graphs with non-trivial structures used to represent certain data domains, such as regions of an image and their neighborhoods. Collective dynamics in a complex network refer to the emergent and interactive behavior of various elements or actors within an interconnected and complex network, where the actions of one element can influence the actions of others. In this work, we propose a semi-supervised image segmentation algorithm that combines the techniques of superpixels, complex networks, and collective dynamics. The method was evaluated under various conditions using the GrabCut database for interactive segmentation. Our method proved to be as good as the best interactive segmentation methods based on graphs and in the precision metric it surpassed all compared methods.

**Keywords:** complex networks; image segmentation; superpixel; collective dynamics; interactive segmentation.

## LISTA DE FIGURAS

Figura 1 – Comparação de tipos de segmentação de imagem: por semântica e instância. . . . .	18
Figura 2 – Segmentação superpixel . . . . .	20
Figura 3 – Visão geral da técnica de segmentação interativa FocalClick. . . . .	21
Figura 4 – Quantidade de publicações relevantes envolvendo GrabCut de 2004 a 2022. . . . .	22
Figura 5 – O usuário marca anotações do fundo e do objeto a ser segmentado, então a segmentação interativa é realizada. . . . .	24
Figura 6 – No aprendizado transdutivo, a inferência em novos exemplos ocorre de maneira direta. . . . .	26
Figura 7 – Em tons de azul é representada a concentração de dados rotulados, em vermelho a função de inferência. No aprendizado transdutivo não há estimação de uma função de inferência. . . . .	27
Figura 8 – Execução do algoritmo SLIC na fotografia de um gato. . . . .	29
Figura 9 – Geração de rede complexa baseado nos superpixels. . . . .	30
Figura 10 – Filtros de Gabor variando a angulação $\theta$ e a convolução resultante. . . . .	32
Figura 11 – Visualização da distância de manhattan e distância euclidiana. . . . .	33
Figura 12 – Rede complexa para execução da dinâmica LCU. Anotação parcial em (a) e resultado final em (b). . . . .	35
Figura 13 – Iteração do sistema LCU em um grafo planar. . . . .	36
Figura 14 – Execução em quatro fases da dinâmica coletiva LCU. . . . .	40
Figura 15 – Construção de superpixels e a rede complexa no modelo EGSIS. . . . .	42
Figura 16 – Etapa do método EGSIS para propagar rótulos iniciais através do sistema dinâmico LCU. . . . .	43
Figura 17 – Máscara de segmentação final e a imagem original. . . . .	43
Figura 18 – Amostra de 16 imagens do <i>dataset</i> GrabCut. . . . .	44
Figura 19 – Exemplo de imagem <i>dataset</i> GrabCut com anotação parcial e a segmentação real ( <i>ground truth</i> ). . . . .	45
Figura 20 – Ilustração das métricas IoU e F1 score. . . . .	47
Figura 21 – Comparação entre diferentes métodos de segmentação interativa no <i>dataset</i> GrabCut. . . . .	50

Figura 22 – Execuções de várias segmentações interativas com EGSIS. A coluna ao lado da imagem mostra a segmentação binária feita pelo modelo, e a próxima coluna mostra a segmentação real. . . . .	51
Figura 23 – Ferramenta para anotação parcial de imagem, com pontos circulares e classes distintas por cor. . . . .	54
Figura 24 – Execução do método EGSIS com a matriz de rótulos gerada pelo anotador. . . . .	55

## LISTA DE TABELAS

Tabela 1 – Regras da anotação Lasso do GrabCut . . . . .	45
Tabela 2 – Resultados dos experimentos ao variar o número de superpixels e o método de extração de características. Em negrito os melhores resultados.	48
Tabela 3 – Resultados comparativos entre o método EGSIS e métodos estado-da-arte para segmentação interativa baseado em grafos. Em negrito os melhores resultados, em vermelho os piores. . . . .	52
Tabela 4 – Resultados do modelo EGSIS com desvio-padrão. . . . .	53

## LISTA DE ALGORITMOS

Algoritmo 1 – SLIC . . . . .	29
Algoritmo 2 – EGSIS . . . . .	41

## LISTA DE ABREVIATURAS E SIGLAS

DNN	<i>Deep Neural Networks</i>
EGSIS	<i>Exploratory Graph-based Semi-supervised Image Segmentation</i>
k-NN	<i>K-nearest neighbours</i>
LCU	<i>Labeled Component Unfolding</i>
SLIC	<i>Simple Linear Iterative Clustering</i>

## LISTA DE SÍMBOLOS

$\mathbf{Y}$	Matriz de rótulos.
$\mathbf{X}$	Matriz de vetores de características.
$\vec{x}$	Vetor de características.
$y$	Rótulo.
$\mathbb{R}^d$	Espaço real com $d$ dimensões.
$\in$	Pertencente.
$\lambda$	Grau de competitividade do LCU entre $[0, 1]$ .
$\tau$	Número de iterações máxima do algoritmo LCU.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>1.1</b>	<b>Trabalhos relacionados</b>	<b>19</b>
<b>1.1.1</b>	<i>GrabCut: Interactive Foreground Extraction Using Iterated Graph Cuts</i>	<b>20</b>
<b>1.1.2</b>	<i>Aplicação de agrupamento semi-supervisionado para segmentação de imagens coloridas</i>	<b>21</b>
<b>1.1.3</b>	<i>FocalClick: Towards Practical Interactive Image Segmentation</i>	<b>21</b>
<b>1.1.4</b>	<i>Interactive image segmentation based on multi-layer random forest classifiers</i>	<b>22</b>
<b>1.2</b>	<b>Justificativas</b>	<b>22</b>
<b>1.3</b>	<b>Objetivo Geral</b>	<b>23</b>
<b>1.4</b>	<b>Objetivos Específicos</b>	<b>23</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>24</b>
<b>2.1</b>	<b>Segmentação interativa de imagens</b>	<b>24</b>
<b>2.2</b>	<b>Aprendizado semi-supervisionado vs. transdução</b>	<b>25</b>
<b>2.3</b>	<b>Superpixels</b>	<b>27</b>
<b>2.3.1</b>	<i>SLIC</i>	<b>28</b>
<b>2.4</b>	<b>Geração de redes complexas</b>	<b>29</b>
<b>2.5</b>	<b>Extração de características</b>	<b>31</b>
<b>2.5.1</b>	<i>Matriz de co-ocorrências</i>	<b>31</b>
<b>2.5.2</b>	<i>Filtros de Gabor</i>	<b>31</b>
<b>2.6</b>	<b>Métricas de similaridade</b>	<b>33</b>
<b>2.6.1</b>	<i>Similaridade euclidiana exponencial</i>	<b>34</b>
<b>2.6.2</b>	<i>Similaridade de Manhattan logarítmica</i>	<b>34</b>
<b>2.7</b>	<b>LCU</b>	<b>35</b>
<b>2.7.1</b>	<i>Modelagem matemática LCU com grafo ponderado</i>	<b>37</b>
<b>2.8</b>	<b>EGSIS</b>	<b>40</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>44</b>
<b>3.1</b>	<b>Métricas de avaliação</b>	<b>46</b>
<b>3.1.1</b>	<i>Precision</i>	<b>46</b>

3.1.2	<i>Recall</i> . . . . .	46
3.1.3	<i>F1 Score</i> . . . . .	46
3.1.4	<i>IoU</i> . . . . .	47
4	<b>RESULTADOS</b> . . . . .	48
4.1	Experimentos com a variação da quantidade de superpixels . . . . .	48
4.2	Avaliação qualitativa . . . . .	49
4.3	Comparação quantitativa com o estado-da-arte da segmentação interativa	52
4.4	Ferramenta de anotação integrada com método EGSIS . . . . .	54
5	<b>CONCLUSÕES E TRABALHOS FUTUROS</b> . . . . .	56
5.1	Limitações e desafios . . . . .	56
5.2	Trabalhos futuros . . . . .	57
	<b>REFERÊNCIAS</b> . . . . .	58

## 1 INTRODUÇÃO

Sistemas de segmentação de imagens têm se tornado populares em variadas aplicações, como, por exemplo, a área de edição de imagem, diagnósticos médicos (ZHOU *et al.*, 2019) e parte da visão computacional necessária pra reconhecimento de objetos, muito importante para desenvolvimento de direção autônoma em carros (FENG *et al.*, 2021). Entre esses motivos e outros, essa área tem uma relevância científica alta considerando a situação social, tecnológica e econômica que é vivida no século XXI.

A segmentação de uma imagem pode ser feita manualmente por um anotador humano marcando as linhas delimitadoras de um objeto. Por outro lado, são conhecidos variados algoritmos para segmentação de imagens baseados em aprendizagem de máquina, que, através de exemplos de segmentação fornecidos para treinamento podem inferir a segmentação de novas imagens.

Na Figura 1 é apresentado um quadro comparativo de operações em uma imagem com balões, incluindo os tipos de segmentação de imagens conhecidos: semântica e instância. Na segmentação semântica o objetivo é segmentar apenas os mesmos tipos de objetos como o mesmo rótulo, na segmentação por instância, cada balão é visto como a mesma classe de balão mas com rótulos associados diferentes.

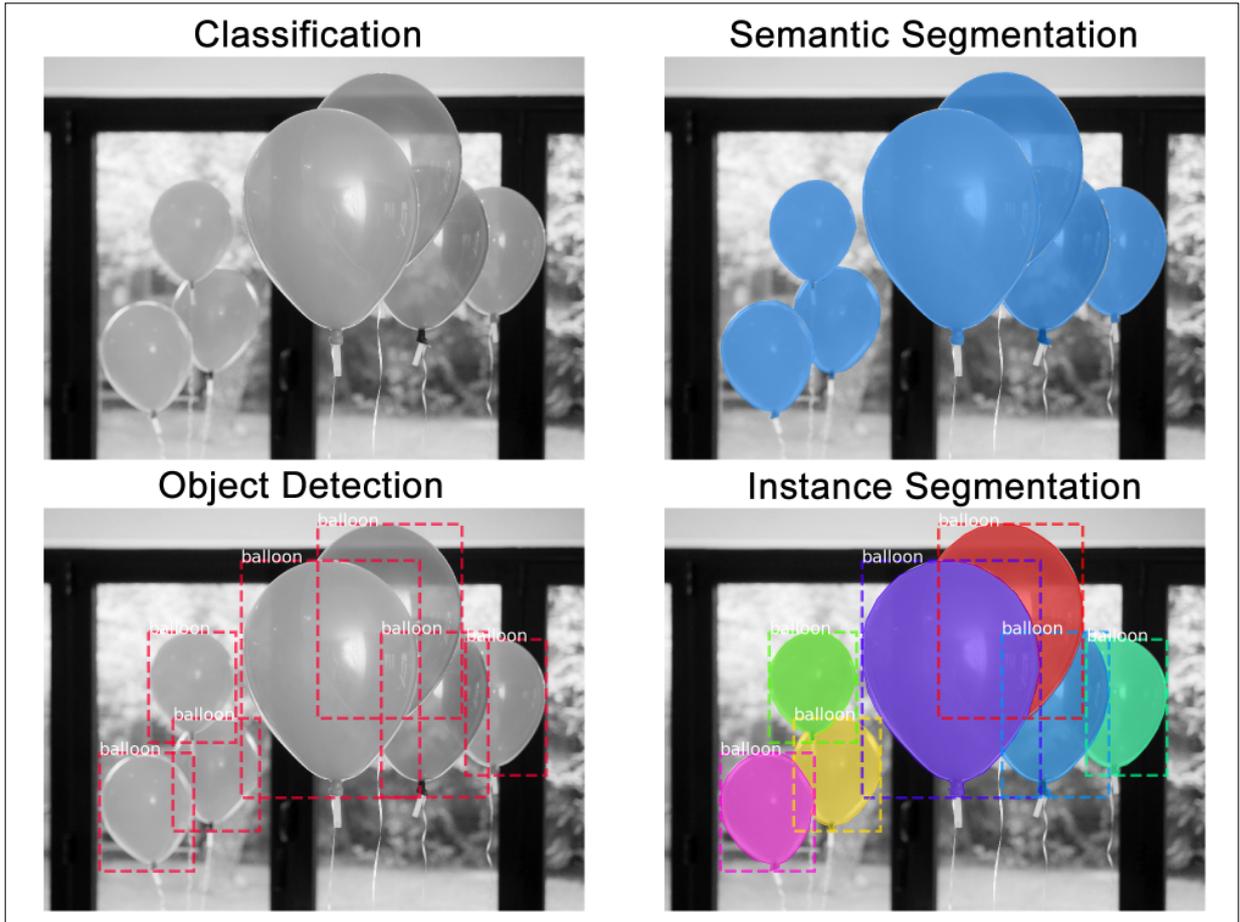
Entre os tipos de aprendizagem de máquina (SAH, 2020), para segmentação de imagens semântica é selecionada neste trabalho especificamente a aprendizagem semi-supervisionada transdutiva<sup>1</sup>. A aprendizagem semi-supervisionada é uma categoria que realiza o aprendizado com poucas rotulações e a maior parte dos dados não são rotulados. Essa categoria se baseia na aprendizagem supervisionada, que possui no treinamento uma base totalmente rotulada, e na aprendizagem não supervisionada, que não possui rótulo algum (exemplo: K-means). Ao considerar a dificuldade de conseguir dados rotulados por humanos em ambientes de uso por especialistas, como imagens médicas e ferramentas de edição de imagem, a abordagem semi-supervisionada se demonstra interessante por precisar de poucos dados rotulados, mas ainda existir uma anotação com viés do especialista interessado (médico, editor).

Os três principais algoritmos clássicos de segmentação de imagem podem ser citados: *Region-Based Segmentation*; *Edge Detection Segmentation*; *Segmentation based on Clustering* (HARALICK; SHAPIRO, 1985). Cada uma dessas técnicas possui limitações

---

<sup>1</sup> mais informações na Seção 2.2

Figura 1 – Comparação de tipos de segmentação de imagem: por semântica e instância.



Fonte: Ahmed (2019)

conhecidas, e entre elas é possível mencionar: haver muitos objetos na imagem pode dificultar a segmentação; tempo computacional elevado; sensibilidade ao contraste em escala cinza.

Um cenário especial de segmentação semântica abordado neste trabalho para aplicação de um esquema de aprendizagem semi-supervisionada é o problema de segmentação interativa<sup>2</sup>, no qual o objetivo é uma segmentação binária que busca segmentar um objeto alvo em relação ao plano de fundo. Para segmentação interativa existem variadas técnicas, com diferentes abordagens, limitações e vantagens (RAMADAN *et al.*, 2020).

Considerando tal situação-problema, este trabalho propõe a construção de uma técnica de segmentação de imagem semi-supervisionada voltada para segmentação interativa utilizando redes complexas e dinâmicas coletivas de tal maneira que possa se equiparar com as melhores técnicas atuais de segmentação interativa baseadas em grafos.

<sup>2</sup> conceito explicado em detalhes na Seção 2.1

## 1.1 Trabalhos relacionados

Técnicas de segmentação de imagens com o paradigma semi-supervisionado estão em foco atualmente no campo médico, como pode ser visto em Luo *et al.* (2021). Nesse artigo, uma das grandes motivações dos autores utilizarem uma técnica semi-supervisionada está relacionada com a dificuldade de adquirir dados rotulados em domínios especializados, como imagens hospitalares.

Em relação ao tópico de redes complexas e dinâmicas coletivas, é possível mencionar o trabalho feito com o algoritmo *Labeled Component Unfolding* (LCU) (VERRI *et al.*, 2018) no qual os principais conceitos sobre resolução de problemas de aprendizagem de máquina semi-supervisionada são explorados em detalhes como uma dinâmica de competição de partículas na relação vértice-arestas de um grafo não-direcionado. Este método de aprendizagem é transdutivo, portanto, difere dos métodos indutivos que estimam uma função de inferência no treinamento, como é o caso de Redes Neurais. Além disso, esse algoritmo tem complexidade computacional linear em relação à quantidade de classes, arestas e vértices.

Para ilustrar uma possível ideia de segmentação de imagens, ao considerar um algoritmo que faça uma transformação do domínio de imagem para um grafo, é possível estabelecer uma relação na qual os vértices representam parte da imagem como um *superpixel* (IBRAHIM; EL-KENAWY, 2020), ou seja, um grupo de subpixels da imagem. Na Figura 2 é apresentado um exemplo de segmentação usando superpixels. Algoritmos de superpixel são não supervisionados em geral, portanto possuem suas limitações quanto ao resultado esperado pelo usuário – logo difícil de ser aceito em aplicações médicas nas quais a opinião do especialista é de alta relevância para o resultado final.

Seguindo essa perspectiva, ao utilizar um algoritmo de extração de *features* de imagens sobre o superpixel, tem-se que o vértice do grafo é neste momento um vetor de características. O sistema de competição proposto no artigo *Network Unfolding Map By Vertex-Edge Dynamics Modeling* (VERRI *et al.*, 2018) pode otimizar o pertencimento de classes (segmentos, nesse caso) baseado na topologia de sua vizinhança e na relação aos vértices conectados. A métrica de similaridade (por exemplo, distância euclidiana, cosseno, etc.) pode ser ajustada de acordo com o problema.

Ao considerar o problema como semi-supervisionado, a pista de ter alguns dos superpixels anotados adicionaria um *bias* parametrizado pelo conhecimento do especialista

Figura 2 – Segmentação superpixel



Fonte: Wang *et al.* (2017)

em uso da ferramenta, como um editor ou um médico. A otimização do pertencimento das classes então seria acionada pela dinâmica coletiva selecionada em questão, que, por acaso, poderia ser o algoritmo LCU mencionado anteriormente.

Por outro lado, ainda há muitas melhorias a serem feitas nessas técnicas, como, por exemplo, analisar as condições de convergência do algoritmo. Isso pode ser um dos resultados deste trabalho, demandando uma análise matemática com auxílio de experimentos.

É importante mencionar que já foi demonstrado em outras situações, como em (RIBAS *et al.*, 2020), que o uso de redes complexas em fusão com redes neurais aleatórias pode gerar um discriminante de textura da imagem de alta relevância como extrator de características. Neste caso, é possível se apoiar nesse resultado como uma evidência de que a investigação de novas técnicas considerando a topologia da imagem através de redes complexas é uma oportunidade de pesquisa.

### ***1.1.1 GrabCut: Interactive Foreground Extraction Using Iterated Graph Cuts***

Neste trabalho, um dos pioneiros em segmentação interativa, os autores Rother *et al.* (2004) desenvolveram uma técnica para segmentação interativa chamada GrabCut baseado em um mapeamento da imagem como um grafo e então um corte entre as arestas é realizado para segmentar o objeto do plano de fundo.

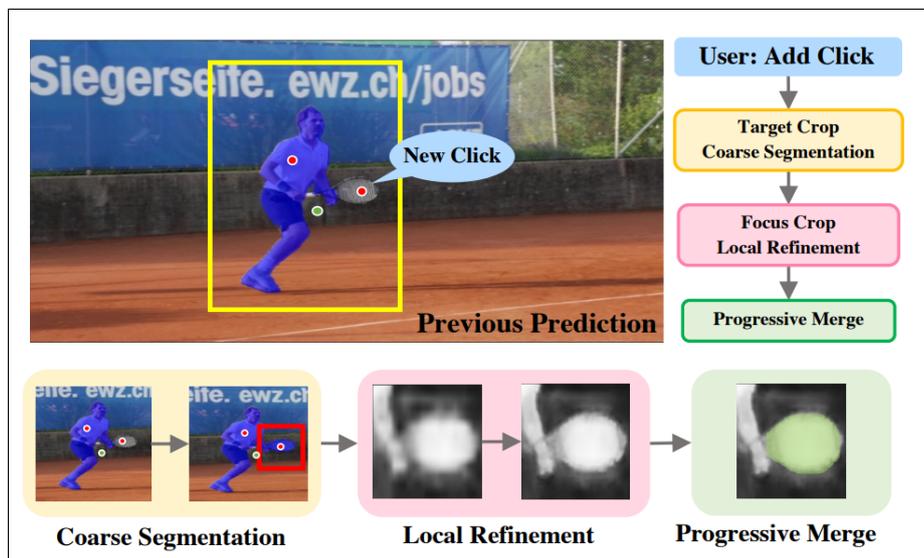
### 1.1.2 Aplicação de agrupamento semi-supervisionado para segmentação de imagens coloridas

Neste trabalho, o autor Lira (2018), na sua tese de graduação, propõe variações de um algoritmo de segmentação de imagem semi-supervisionado combinando algoritmos de agrupamento, como *Fuzzy C-Means*, Algoritmo de Pedrycs, Algoritmo Semi-supervisionado Padrão (sSSC) e Algoritmo Semi-supervisionado Regularizado por Entropia (ESSC).

### 1.1.3 FocalClick: Towards Practical Interactive Image Segmentation

Neste trabalho, os autores Chen *et al.* (2022) criam uma técnica de segmentação interativa em busca da praticidade, ao medir dois aspectos importantes além de qualidade de segmentação: necessidade de anotação e tempo de execução. O trabalho em sua metodologia de avaliação utiliza métricas para minimizar interações que os usuários tenham que fazer para alcançar uma segmentação de qualidade. Essa técnica se baseia em um algoritmo iterativo que possui capacidades de correção ao incluir novas marcações. Na Figura 3 é possível visualizar uma visão geral da técnica:

Figura 3 – Visão geral da técnica de segmentação interativa FocalClick.



Fonte: Chen *et al.* (2022)

### 1.1.4 *Interactive image segmentation based on multi-layer random forest classifiers*

Neste trabalho, as autoras Shan *et al.* (2023) criam uma nova técnica para segmentação interativa de imagens combinando uma pré-segmentação usando superpixels e duas camadas de modelos *random forest* para classificar o agrupamento entre os superpixels numa máscara de segmentação resultante.

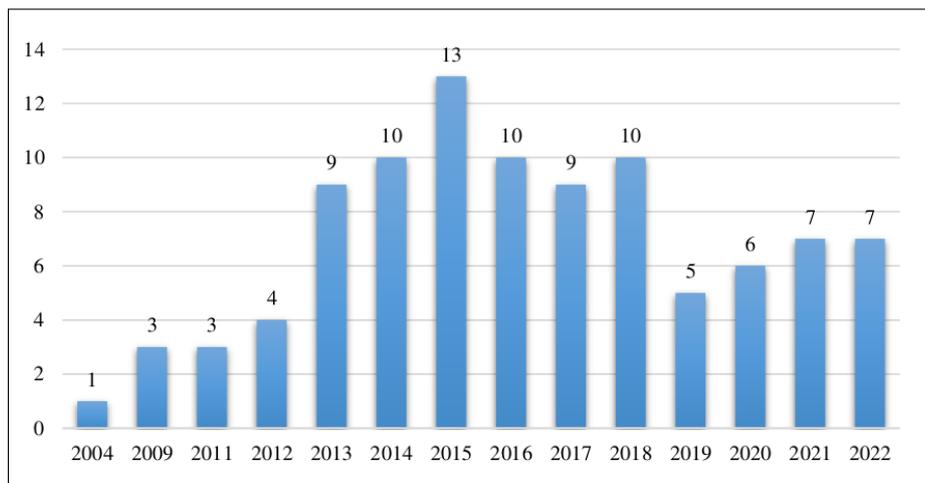
## 1.2 Justificativas

Como mencionado na Introdução, os algoritmos conhecidos de segmentação de imagens possuem restrições pertinentes que podem dificultar o uso das técnicas em alguns problemas, como imagens médicas e edição de imagens.

Esses problemas são endereçados ao desenvolvimento de uma nova técnica de segmentação de imagens que explora outros novos ramos de aprendizagem de máquina semi-supervisionada além das *Deep Neural Networks* (DNN), neste caso utilizando redes complexas e dinâmicas coletivas.

O campo de pesquisa de segmentação interativa também tem tido aumento na publicação de artigos de relevância, em especial a publicação de artigos com a palavra-chave GrabCut (tipo de algoritmo para segmentação interativa), como pode ser visto na Figura 4:

Figura 4 – Quantidade de publicações relevantes envolvendo GrabCut de 2004 a 2022.



Fonte: Wang *et al.* (2023)

### 1.3 Objetivo Geral

Desenvolver uma nova técnica de segmentação de imagens semi-supervisionada que possa ser equiparável ao estado-da-arte, com foco em segmentação interativa.

### 1.4 Objetivos Específicos

- Explorar técnicas de redes complexas e dinâmicas coletivas sobre o problema de segmentação de imagens.
- Aplicar em casos variados de segmentação de imagens, como objetos comuns, carros, pessoas, paisagens, etc.
- Avaliar o impacto da segmentação por superpixel na segmentação final.
- Desenvolver uma ferramenta para segmentação interativa utilizando o método de segmentação proposto.

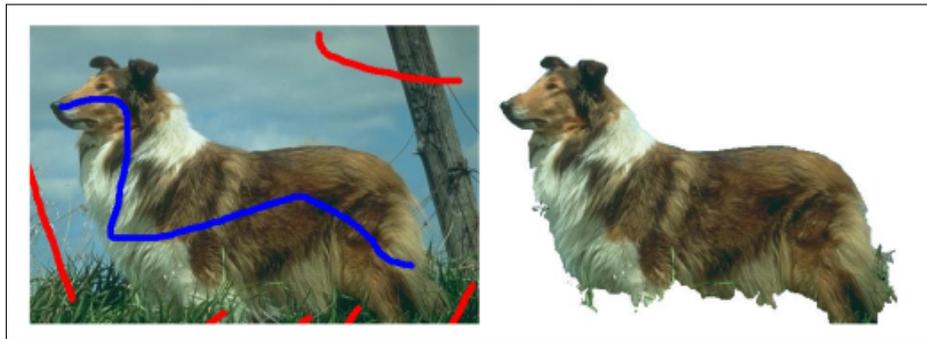
## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados vários conceitos e técnicas para fundamentar o desenvolvimento do método proposto, descrito na Seção 2.8.

### 2.1 Segmentação interativa de imagens

A segmentação interativa de imagens é um processo de divisão de uma imagem digital em várias partes ou regiões, geralmente para tornar a imagem mais fácil de analisar e processar (RAMADAN *et al.*, 2020). Este processo é ‘interativo’ porque envolve a entrada do usuário para ajudar a orientar ou refinar o processo de segmentação.

Figura 5 – O usuário marca anotações do fundo e do objeto a ser segmentado, então a segmentação interativa é realizada.



Fonte: Duchenne *et al.* (2008)

De acordo com a Figura 5, na segmentação interativa de imagens, o usuário pode selecionar regiões de interesse, definir marcadores ou fazer anotações na imagem. Essas entradas do usuário são então usadas para informar o algoritmo de segmentação sobre como dividir a imagem. Por exemplo, o usuário pode desenhar uma linha ao redor de um objeto de interesse, e o algoritmo de segmentação irá então tentar dividir a imagem de tal forma que o objeto de interesse seja isolado em sua própria região.

Este tipo de segmentação de imagem é útil em uma variedade de aplicações, incluindo processamento de imagens médicas, visão computacional, reconhecimento de padrões e muitos outros campos nos quais é útil dividir uma imagem em regiões distintas com base em critérios definidos pelo usuário.

## 2.2 Aprendizado semi-supervisionado vs. transdução

Existem três principais categorias de aprendizado de máquina: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semi-supervisionado. No aprendizado supervisionado, durante a etapa de treinamento existe uma base de dados totalmente rotulada, no aprendizado não supervisionado não é disponibilizado nenhum rotulamento dos dados. Enquanto isso, o aprendizado semi-supervisionado está entre essas duas categorias.

O aprendizado semi-supervisionado é um método de aprendizado de máquina que envolve o uso de um grande volume de dados não rotulados e um pequeno volume de dados rotulados para treinar modelos de aprendizado de máquina.

Formalmente, pode-se definir o aprendizado semi-supervisionado da seguinte maneira:

Dado um conjunto de dados de treinamento  $\mathbf{X} = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ , tal que os vetores de características tenham dimensão  $d$ , isto é,  $\vec{x}_n \in \mathbb{R}^d$ , onde apenas um subconjunto  $\vec{Y} = \{y_1, y_2, \dots, y_m\}$  em que  $(m < n)$  tem rótulos correspondentes, o objetivo do aprendizado semi-supervisionado é usar tanto o conjunto de dados rotulado quanto o não rotulado para aprender a função  $f : \mathbf{X} \rightarrow \vec{Y}$  que pode prever o rótulo  $y$  para um novo exemplo  $\vec{x}$ .

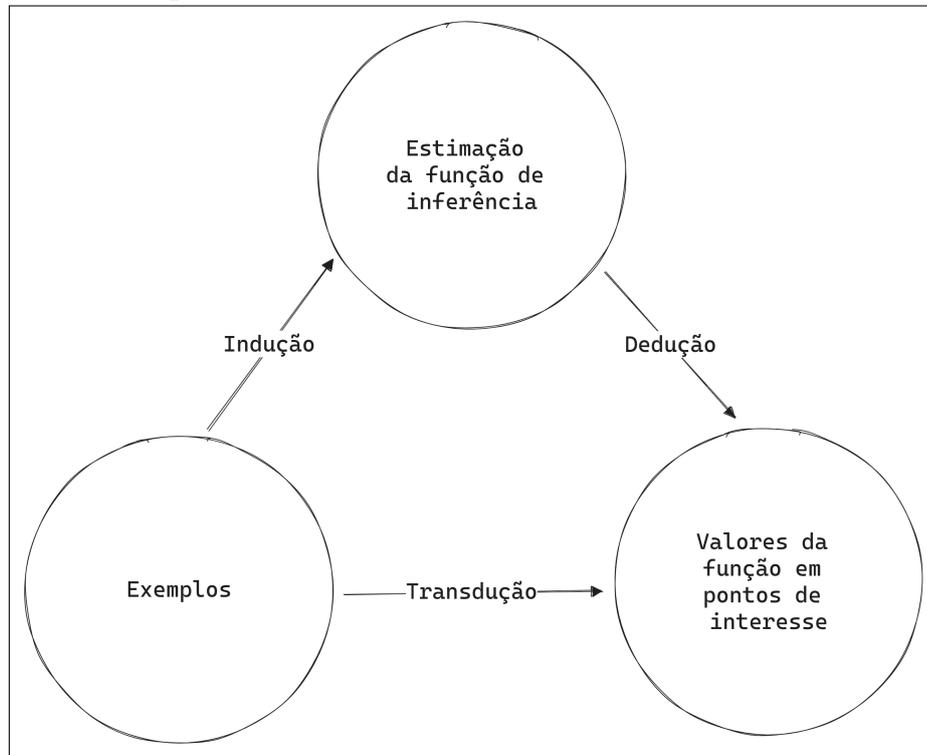
O aprendizado semi-supervisionado é baseado na suposição de que os dados não rotulados podem fornecer informações adicionais que podem ser usadas para melhorar a precisão do modelo de aprendizado de máquina. Isso é feito através de várias técnicas, como a propagação de rótulos, na qual os rótulos são propagados dos dados rotulados para os dados não rotulados, ou a aprendizagem auto-supervisionada, na qual o modelo é treinado para prever partes dos dados a partir de outras partes.

Por outro lado, existe um diferente tipo de aprendizado semi-supervisionado que não realiza a etapa de estimar a função  $f$ . Algoritmos que estimam essa função são descritos como indutivos, pois após o treinamento, para classificar novos dados, realizam uma inferência por indução ao aplicar a função estimada.

Em contraponto, existem algoritmos que não estimam tal função e apenas realizam a inferência direta entre os dados rotulados disponíveis e os não rotulados. Isso é chamado de transdução e um algoritmo bem conhecido com essa característica é o

classificador *K-nearest neighbours* (k-NN). Na Figura 6 são ilustradas as diferenças entre transdução e indução no processo de aprendizagem:

Figura 6 – No aprendizado transdutivo, a inferência em novos exemplos ocorre de maneira direta.



Fonte: Adaptada de Vapnik (1995)

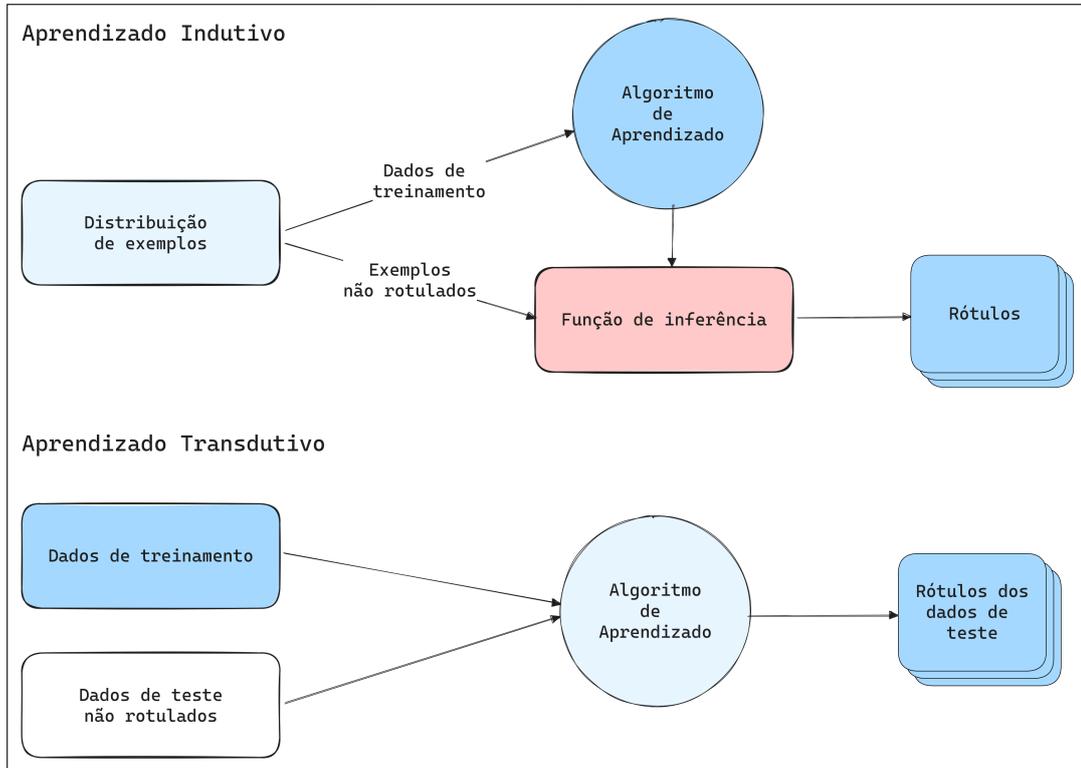
No texto de Vapnik (2006), o criador do famoso algoritmo SVM estabelece uma profunda formalização dos problemas de aprendizado semi-supervisionado e inferência transdutiva. No final do texto, ele declara algumas reflexões sobre a solução de problemas em aprendizagem de máquina. Em uma delas, ele diz que, ao resolver um problema, não tente resolver o problema geral como um passo intermediário, tente obter a resposta que você precisa, mas não a mais geral. Por fim, Vapnik deixa uma sugestão relevante:

“Do not estimate a function if you need to estimate values at given points. (*Try to perform transduction, not induction.*)” (VAPNIK, 2006, p. 472)

Em tradução-livre: “Não estime uma função se você precisa estimar valores em dados pontos. (Tente executar transdução, não indução)”. Essa sugestão é uma evidência que reforça a motivação deste trabalho. Ao desenvolver um algoritmo transdutivo para segmentação de imagens de maneira assistida, não é estimada uma função geral como um passo intermediário, mas a segmentação é realizada diretamente baseada nas rotulações

iniciais que o usuário forneceu. Na Figura 7 está contida uma ilustração comparativa entre aprendizado indutivo e transdutivo.

Figura 7 – Em tons de azul é representada a concentração de dados rotulados, em vermelho a função de inferência. No aprendizado transdutivo não há estimação de uma função de inferência.



Fonte: Adaptada de Vapnik (1995)

## 2.3 Superpixels

Superpixels fazem parte de um grupo de algoritmos de clusterização em processamento de imagens que ganhou popularidade nos últimos anos na comunidade de visão computacional (IBRAHIM; EL-KENAWY, 2020). Em vez de processar uma imagem pixel por pixel, agrupam-se pixels vizinhos semelhantes em uma entidade maior, conhecida como superpixel.

O conceito de superpixels foi introduzido para superar as limitações do processamento pixel a pixel, que não leva em consideração a estrutura global da imagem. Os superpixels, por outro lado, mantêm a estrutura da imagem e reduzem a complexidade do processamento de imagens, tornando-o mais eficiente.

Os superpixels são formados com base na similaridade dos pixels em termos de

cor, intensidade e localização na imagem. Eles são usados em uma variedade de aplicações, incluindo segmentação de imagem, rastreamento de objetos, reconhecimento de objetos, entre outros.

Em resumo, os superpixels são uma técnica eficaz para simplificar a representação de uma imagem e aumentar a eficiência do processamento de imagens, mantendo a informação visual importante.

Atualmente, já é possível encontrar muitas técnicas baseadas em superpixels com diferentes características, complexidades computacionais, métodos e eficiência. No artigo de Wang *et al.* (2017), é realizado um *benchmark* com 15 algoritmos do tipo superpixel categorizados em três grupos: baseado em grafos, baseado em otimização de gradiente e baseado em análise de textura. Neste trabalho, é selecionado um dos mais simples: *Simple Linear Iterative Clustering* (SLIC).

### 2.3.1 SLIC

O algoritmo SLIC (ACHANTA *et al.*, 2010; ACHANTA *et al.*, 2012) é um método para segmentação de imagens baseado em superpixel, é um dos mais simples e pode ser visto como uma variação do algoritmo de clusterização K-means expandindo o espaço euclidiano ao incluir também o espaço de cores. Ele divide uma imagem em segmentos menores, chamados superpixels, que compartilham características semelhantes, como cor e textura.

O algoritmo SLIC possui três hiperparâmetros: *segments*, *sigma* e *compactness*.

1. **Segments:** Número de superpixels que será usado para segmentar a imagem.
2. **Sigma:** Este parâmetro é usado para suavizar a imagem antes da segmentação. Um valor maior de sigma resultará em uma imagem mais suave, o que pode ajudar a reduzir o ruído e os detalhes finos. No entanto, um valor muito alto pode resultar em perda de detalhes importantes.
3. **Compactness** Este parâmetro controla o equilíbrio entre a coerência de cor e a proximidade espacial na formação de superpixels. Um valor maior de compactness fará com que os superpixels sejam mais quadrados, enquanto um valor menor fará com que os superpixels sigam mais de perto os limites da imagem. Portanto, a compactness pode ser ajustada para obter superpixels que são mais representativos da estrutura da imagem.

Uma explicação passo-a-passo de como o SLIC funciona pode ser entendida dessa maneira:

---

**Algoritmo 1: SLIC**

---

**Entrada:** Imagem a ser segmentada

**Resultado:** Matriz de superpixels

**início**

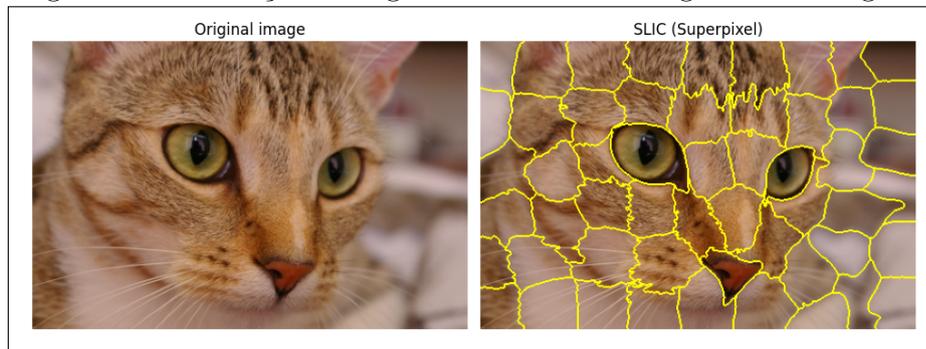
1. **Inicialização:** O algoritmo começa selecionando alguns pixels na imagem como centros de superpixels. Esses centros são espaçados uniformemente pela imagem;
2. **Atribuição:** Em seguida, para cada pixel na imagem, o algoritmo calcula a distância entre esse pixel e todos os centros de superpixels. A distância é calculada com base na cor (ou intensidade de níveis de cinza) e na proximidade espacial. O pixel é então atribuído ao superpixel cujo centro está mais próximo;
3. **Atualização:** Depois que todos os pixels foram atribuídos a um superpixel, o algoritmo recalcula os centros de superpixels como a média de todos os pixels dentro de cada superpixel;
4. O processo de atribuição e atualização é repetido várias vezes até que o algoritmo alcance a condição de convergência, ou seja, até que os centros de superpixels parem de mudar significativamente;
5. O resultado final é uma segmentação da imagem em superpixels, na qual cada superpixel é um grupo de pixels com características semelhantes.

**fim**

---

Na Figura 8, uma execução do SLIC é possível de ser visualizada.

Figura 8 – Execução do algoritmo SLIC na fotografia de um gato.



Fonte: Elaborado pelo autor baseado nas imagens disponíveis em Walt *et al.* (2014).

## 2.4 Geração de redes complexas

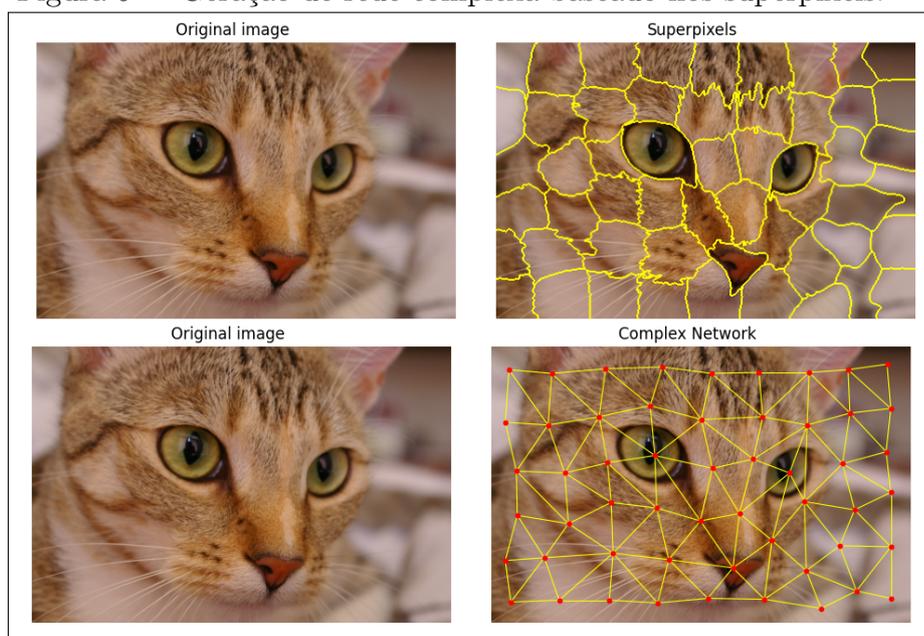
Redes complexas são grafos de alta complexidade. Existem variados algoritmos para geração de redes complexas (COSTA *et al.*, 2007). Redes complexas podem ser usadas como um domínio de dados para realizar tarefas como classificação de imagens (LIMA *et al.*, 2015), segmentação de imagens, identificação de comunidades e também extração de características (RIBAS *et al.*, 2020).

Neste trabalho, o uso de redes complexas é realizado de uma maneira acoplada ao algoritmo de clusterização inicial da imagem (superpixel). Nesse cenário, cada superpixel gerado na imagem é um vértice e as arestas são geradas com base na vizinhança.

Apesar de considerar o tema como redes complexas, esse cenário em particular gera um grafo planar pela maneira como as arestas são criadas. Por outro lado, seria possível também modificar a geração da rede complexa para considerar as arestas do grafo baseado num raio parametrizado de superpixel em relação a um *threshrod* e selecionar as k-NN similares como arestas válidas.

Essa etapa de geração da rede complexa é crucial para a execução da dinâmica coletiva explorada na Seção 2.7, que é um dos pontos centrais deste trabalho. Na Figura 9, um exemplo de geração de rede complexa é apresentado conectado com a Figura 8, ao executar o algoritmo SLIC.

Figura 9 – Geração de rede complexa baseado nos superpixels.



Fonte: Elaborado pelo autor baseado nas imagens disponíveis em Walt *et al.* (2014).

## 2.5 Extração de características

### 2.5.1 Matriz de co-ocorrências

O método de extração de características da matriz de co-ocorrências (HARALICK, 1979) é uma técnica utilizada em processamento de imagem e visão computacional para extrair características texturais de uma imagem.

Formalmente, uma matriz de co-ocorrência  $C$  é definida sobre uma imagem  $I$ , para um deslocamento  $\Delta x, \Delta y$ , como:

$$C_{\Delta x, \Delta y}(i, j) = \sum_{x=1}^n \sum_{y=1}^m \begin{cases} 1, & \text{if } I(x, y) = i \text{ e } I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{caso contrário.} \end{cases} \quad (2.1)$$

Na Equação 2.1,  $C_{\Delta x, \Delta y}(i, j)$  é o número de vezes que o par de pixels com intensidades  $i$  e  $j$  ocorre em dois pixels separados por determinadas distância e orientação na imagem  $I$ .

A matriz de co-ocorrência é tipicamente normalizada dividindo cada elemento pelo número total de pares de pixels na imagem, resultando em uma matriz de probabilidade de co-ocorrência.

A partir desta matriz, várias características texturais podem ser extraídas, como contraste, correlação, energia e homogeneidade. Estas características podem ser usadas para tarefas como classificação de textura, segmentação de imagem, entre outras. Após a aplicação do método em cada canal RGB da imagem, são extraídas cinco estatísticas: média, mediana, variância, desvio-padrão e 25-quantis. Então as características são concatenadas em um vetor de características de 15 dimensões.

### 2.5.2 Filtros de Gabor

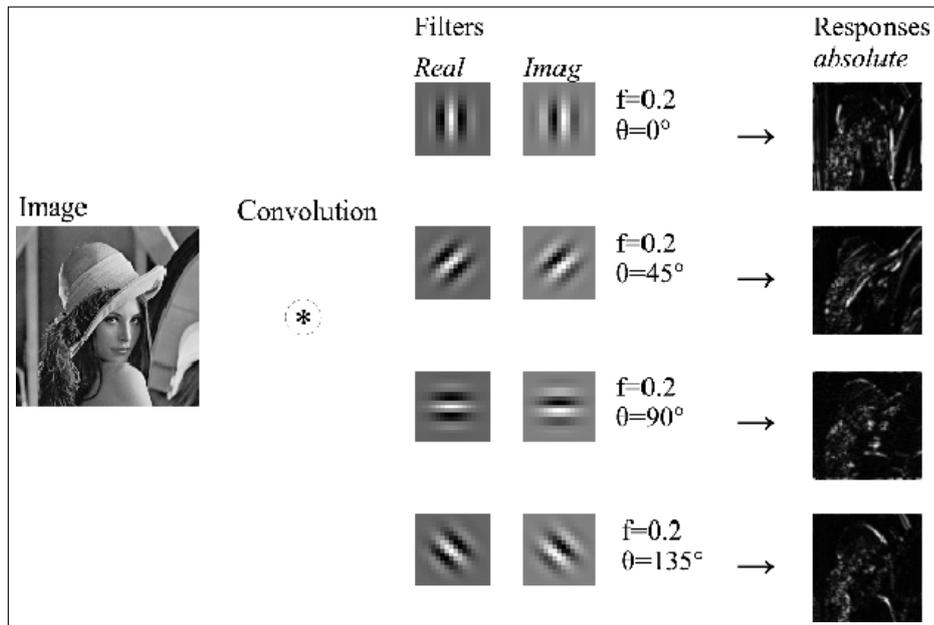
Os filtros de Gabor (DAUGMAN, 1988) são uma família de filtros de convolução usados principalmente na análise de texturas e na visão computacional para a extração de características. Eles são nomeados em homenagem a Dennis Gabor, o físico que ganhou o Prêmio Nobel por sua invenção e desenvolvimento do método holográfico.

Os filtros de Gabor são essencialmente filtros passa-banda orientados e localizados no espaço de frequência. Eles são muito úteis para a extração de características

porque têm propriedades ótimas de localização no espaço e na frequência, o que os torna muito adequados para a captura de características locais e direcionais em uma imagem.

Os filtros de Gabor são definidos por duas funções: uma função gaussiana (para o domínio espacial) e uma função harmônica complexa (para o domínio da frequência). A função gaussiana serve para limitar a aplicação do filtro a uma pequena região da imagem (localização espacial), enquanto a função harmônica complexa permite capturar características em uma determinada orientação. Na Figura 10 são ilustrados variações dos filtros de Gabor com diferentes parâmetros e a convolução resultante.

Figura 10 – Filtros de Gabor variando a angulação  $\theta$  e a convolução resultante.



Fonte: Kämäräinen (2012)

A operação de convolução entre uma imagem qualquer  $x$  e um filtro de Gabor  $k$ , pode ser dada pela seguinte equação:

$$(y * k)(i, j) = \sum_m \sum_n x(i - m, j - n)k(m, n). \quad (2.2)$$

A extração de características usando filtros de Gabor geralmente envolve a convolução da imagem com um conjunto de filtros de Gabor. Cada filtro é sensível a características de uma determinada escala e orientação. A resposta do filtro pode então ser usada como uma representação da presença de tais características na imagem.

Neste trabalho, são gerados 16 filtros de Gabor que foram avaliados experi-

mentalmente com variados parâmetros. Após a aplicação do filtro em cada canal RGB da imagem, são extraídos cinco estatísticas: média, mediana, variância, desvio-padrão e 25-quantis. Com isso, todas as cinco estatísticas dos 16 filtros aplicados em cada um dos três canais são concatenadas num único vetor de características resultante com 240 ( $16 \times 5 \times 3$ ) dimensões.

## 2.6 Métricas de similaridade

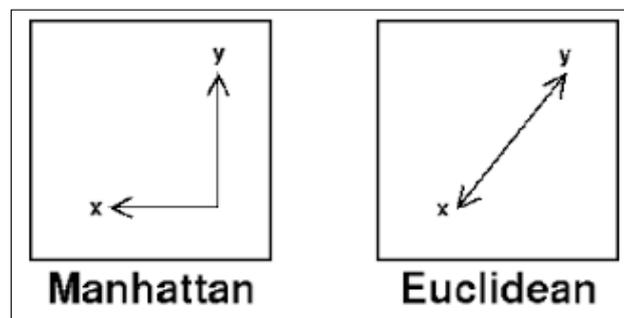
Métricas de similaridade neste trabalho são calculadas com o propósito de comparar a similaridade entre os vetores de características dos superpixels. Para as técnicas de extração de características descritas na Seção 2.5, cada técnica se comportou melhor com uma métrica de similaridade diferente.

Uma função de similaridade entre dois vetores de características retorna um valor maior quando os vetores são similares e quanto mais diferentes sejam, mais a similaridade tende a zero.

A estrutura de uma métrica de similaridade pode ser criada a partir de uma métrica de distância. Neste trabalho, são exploradas duas métricas de similaridades a partir das seguintes métricas de distância: distância euclidiana e distância de manhattan.

A distância euclidiana é o módulo da reta que conecta dois pontos no espaço. A distância de manhattan é a soma da diferença absoluta entre dois pontos para cada dimensão. Na Figura 11 é ilustrada a disposição de dois pontos  $x$  e  $y$  no espaço euclidiano com os módulos de segmento de reta usados para compor a distância de cada métrica.

Figura 11 – Visualização da distância de manhattan e distância euclidiana.



Fonte: Singh (2019)

Para as equações das distâncias, tem-se a seguinte equação para a distância euclidiana:

$$ed(\vec{x}, \vec{y}) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2}. \quad (2.3)$$

Para a distância de manhattan, tem-se a seguinte equação:

$$md(\vec{x}, \vec{y}) = |y_1 - x_1| + |y_2 - x_2| + \cdots + |y_n - x_n|. \quad (2.4)$$

### 2.6.1 Similaridade euclidiana exponencial

Para a técnica de extração de característica de matrizes de co-ocorrência, detalhada na Seção 2.5.1, a métrica de similaridade que obteve os melhores resultados foi a distância euclidiana pelo inverso da exponencial. A equação é definida como:

$$ed\_sim(\vec{x}, \vec{y}) = \frac{1}{e^{ed(\vec{x}, \vec{y})}}. \quad (2.5)$$

### 2.6.2 Similaridade de Manhattan logarítmica

Para a técnica envolvendo filtros de Gabor, a distância entre os vetores de características tiveram valores tão altos que, ao usar a similaridade euclidiana (exponencial ou não), a maior parte dos valores chegavam a zero. Por esse motivo, a conversão para o domínio logarítmico foi usada para suavizar o crescimento exagerado dos valores, obtendo bons resultados. A definição para essa equação de similaridade é:

$$md\_sim(\vec{x}, \vec{y}) = \frac{1}{1 + \ln(1 + md(\vec{x}, \vec{y}))}. \quad (2.6)$$

Na Equação 2.6, os valores 1 adicionados ao denominador foram necessários para evitar uma possível divisão por zero e logaritmo de zero, ambas operações que são indefinidas na matemática.

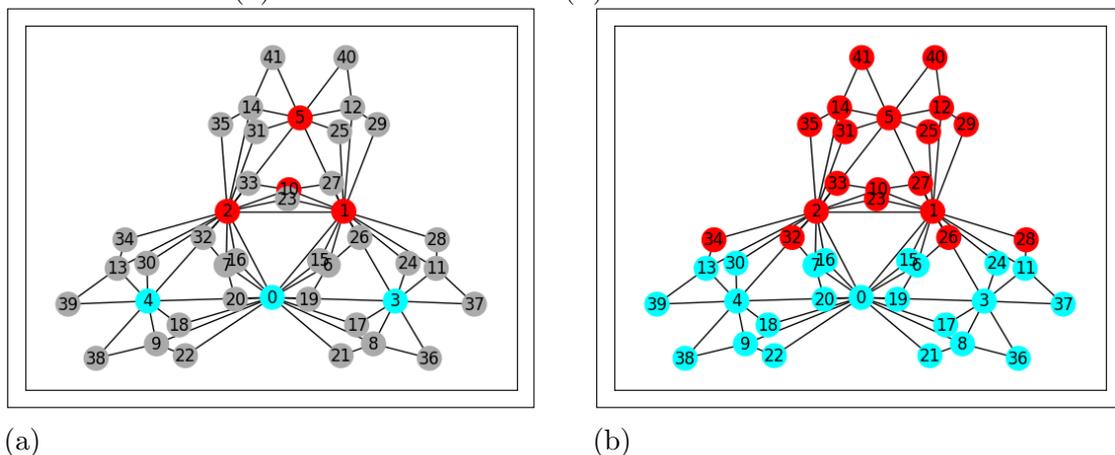
## 2.7 LCU

O algoritmo LCU, que foi desenvolvido por um brasileiro (VERRI *et al.*, 2018), é uma dinâmica coletiva baseada em propagação de rótulos numa rede complexa. Essa dinâmica coletiva é modelada como um sistema dinâmico com geração de partículas de rotulação nas suas fontes (os vértices rotulados).

Esse algoritmo possui critérios de sobrevivência das partículas inspirado em comportamentos da natureza e sociedade. Por exemplo, o hiperparâmetro  $\lambda$  do algoritmo denota um fator de competição entre 0 e 1 sobre a dificuldade que as partículas terão ao trafegar nos vértices, tornando a sobrevivência delas mais difícil ao percorrer o grafo. Para esse parâmetro, 0 significa um passeio aleatório no grafo, 1 significa máxima competitividade. Embora as partículas nasçam nos vértices, a competição por dominação acontece nas arestas. No final, o vértice será marcado com o novo rótulo baseado no tipo de partícula que mais conseguiu dominar arestas desse vértice.

Nesse caso particular, a implementação proposta neste trabalho é ligeiramente diferente da proposta originalmente, pois na rede complexa no cenário que o autor propõe as arestas não possuem peso. Uma modificação é realizada para incluir a similaridade de imagem entre dois superpixels, dessa maneira é criado um fator de aumento da probabilidade das partículas visitarem os nós mais promissores (texturas mais similares). Como discutido na Seção 2.2, esse cenário de aprendizado é semi-supervisionado transdutivo: poucos rótulos estão disponíveis e nenhuma função de inferência é estimada. A inferência acontece diretamente entre os pontos rotulados e os não rotulados.

Figura 12 – Rede complexa para execução da dinâmica LCU. Anotação parcial em (a) e resultado final em (b).

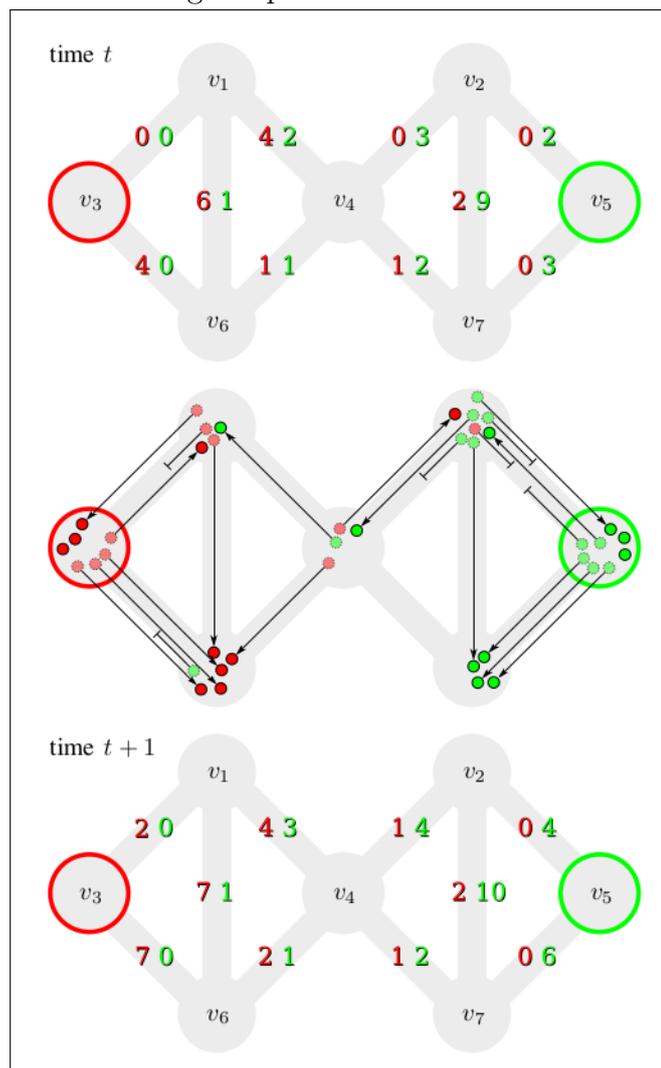


Fonte: Elaborado pelo autor.

Na Figura 12, é possível ver a execução do algoritmo numa rede complexa densa aleatória, no início há alguns poucos vértices anotados em azul e vermelho. No final da execução os rótulos são propagados semanticamente baseado na disputa entre as partículas para dominar arestas, e, por fim, são marcados os vértices dominados por aquela classe.

Em mais detalhes, durante uma única iteração do sistema na Figura 13, os vértices  $v_3$  e  $v_5$  são representados como fontes geradoras de partículas vermelhas e verdes, respectivamente. O passeio dessas partículas entre os vértices é ilustrado por uma seta, que quando interrompida significa que a partícula morreu antes de alcançar o próximo vértice. Algumas partículas não sobreviveram ao processo, outras se mantiveram no sistema. Na última imagem, é possível ver uma contagem do número de partículas por classe que passaram pelas arestas.

Figura 13 – Iteração do sistema LCU em um grafo planar.



Fonte: Verri *et al.* (2018)

### 2.7.1 Modelagem matemática LCU com grafo ponderado

Considera-se uma rede complexa expressada por um simples grafo ponderado não-direcionado  $G = (\mathcal{V}, \mathcal{E})$ , em que  $\mathcal{V}$  é o conjunto de vértices e  $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$  é o conjunto de arestas. A rede possui  $|\mathcal{V}| = l + u$  vértices que podem ser representações de dados rotulados ( $l$ ) ou não rotulados ( $u$ ). Os vértices anotados podem ter um rótulo  $y_i = \{1, \dots, C\}$ , em que  $C$  é o número total de classes, além disso possuem um vetor de características  $\vec{x}_i$  da representação daquele dado. A rede é representada por uma matriz de adjacência  $A = (a_{ij})$ , em que  $a_{ij} = a_{ji} = \text{similarity}(\vec{x}_i, \vec{x}_j)$ , na qual as arestas possuem um peso que se refere à similaridade entre os vetores de características representados pelos vértices. Nessa situação, supõe-se que  $l \ll u$ , ou seja, há muito mais dados não rotulados do que rotulados, na qual caracteriza-se um cenário para aprendizado semi-supervisionado.

Os vértices rotulados são chamados de fontes, que geram partículas com a classe que responde ao vértice. Partículas que alcançam um vértice fonte com classe diferente são eliminadas automaticamente do sistema.

O sistema dinâmico não-linear  $X(t)$  que representa o LCU é definido matematicamente como:

$$X(t) := \begin{bmatrix} \mathbf{n}^c(t) = [n_i^c(t)]_i \\ N^c(t) = (n_{ij}^c(t))_{i,j} \\ \Delta^c(t) = (\delta_{ij}^c(t))_{i,j} \end{bmatrix}, \quad (2.7)$$

em que  $\mathbf{n}^c(t)$  é um vetor linha cujos elementos  $n_i^c(t)$  representam a população de partículas com rótulo  $c$  em cada vértice  $v_i$  no tempo  $t$ . Os elementos  $n_{ij}^c(t)$  representam o número de partículas da classe  $c$  que se movem do vértice  $v_i$  para  $v_j$  no tempo  $t$ , enquanto  $\delta_{ij}^c(t)$  representa a movimentação acumulada de partículas da classe  $c$  entre a aresta  $(i, j)$  no tempo  $t$ .

A função determinística de evolução do sistema dinâmico é definida por:

$$\phi : \begin{cases} \mathbf{n}^c(t+1) = P^c(X(t)) \times \mathbf{n}^c(t) + \mathbf{g}^c(X(t)) \\ N^c(t+1) = \text{diag } \mathbf{n}^c(t) \times P^c(X(t)) \\ \Delta^c(t+1) = \Delta^c(t) + N^c(t+1). \end{cases} \quad (2.8)$$

Entre as funções ainda não definidas na função de evolução, destaca-se a de

probabilidade de sobrevivência das partículas  $P^c(X(t))$ , na qual os elementos dessa matriz são definidos como:

$$p_{ij}^c(X(t)) := \begin{cases} 0 & \text{se } v_j \in \mathcal{L} \text{ e } y_j \neq c, \\ \frac{a_{ij}}{\deg v_i} \left(1 - \lambda \sigma_{ij}^c(X(t))\right) & \text{do contrário,} \end{cases} \quad (2.9)$$

em que  $\deg v_i$  representa o grau do vértice (número de arestas conectadas) e a variável  $\sigma_{ij}^c$  é definida como:

$$\sigma_{ij}^c(X(t)) := \begin{cases} 1 - \frac{n_{ij}^c(t) + n_{ji}^c(t)}{S} & \text{se } S > 0, \\ 1 - \frac{1}{C} & \text{do contrário,} \end{cases} \quad (2.10)$$

sendo  $S = \sum_{q=1}^C n_{ij}^q(t) + n_{ji}^q(t)$ .

A modificação dessa dinâmica coletiva em relação à proposta original em (VERRI *et al.*, 2018) é realizada em  $a_{ij}$ , enquanto no trabalho original o grafo é não-ponderado e, portanto, todas as arestas possuem valor unitário. Neste trabalho a aresta representa um valor de similaridade entre os vetores de características anexados ao vértice  $v_i$  e  $v_j$ , dessa maneira aumentando a probabilidade de sobrevivência de partículas que se movimentam por arestas com maior similaridade.

A última função  $\mathbf{g}^c(X(t))$  representa a geração de partículas nos vértices fontes no tempo  $t$ , e é definida como:

$$g_i^c(X(t)) := \rho_i^c \max \left\{ 0, \sum_{j=1}^{|\mathcal{V}|} (\mathbf{n}_j^c(0) - \mathbf{n}_j^c(t)) \right\},$$

$$\rho_i^c := \begin{cases} \frac{\deg v_i}{\sum_{v_j \in \mathcal{G}^c} \deg v_j} & \text{se } v_i \in \mathcal{G}^c, \\ 0 & \text{do contrário.} \end{cases} \quad (2.11)$$

Na Equação 2.11,  $\mathcal{G}^c = \{v_i | v_i \in \mathcal{L}, y_i = c\}$  é o conjunto de fontes de partículas que pertencem à classe  $c$  e  $\mathcal{L}$  é o conjunto dos vértices rotulados.

Após alcançar a iteração máxima de um valor escolhido para  $\tau$ , é possível dividir a rede complexa em sub-grafos, primeiramente agrupando as arestas por maior dominação de classe:

$$\mathcal{E}^c(t) := \left\{ (i, j) \mid \arg \max_q \left( \delta_{ij}^q(t) + \delta_{ji}^q(t) \right) = c \right\}. \quad (2.12)$$

E, então, define-se o grafo por classe  $c$  como:

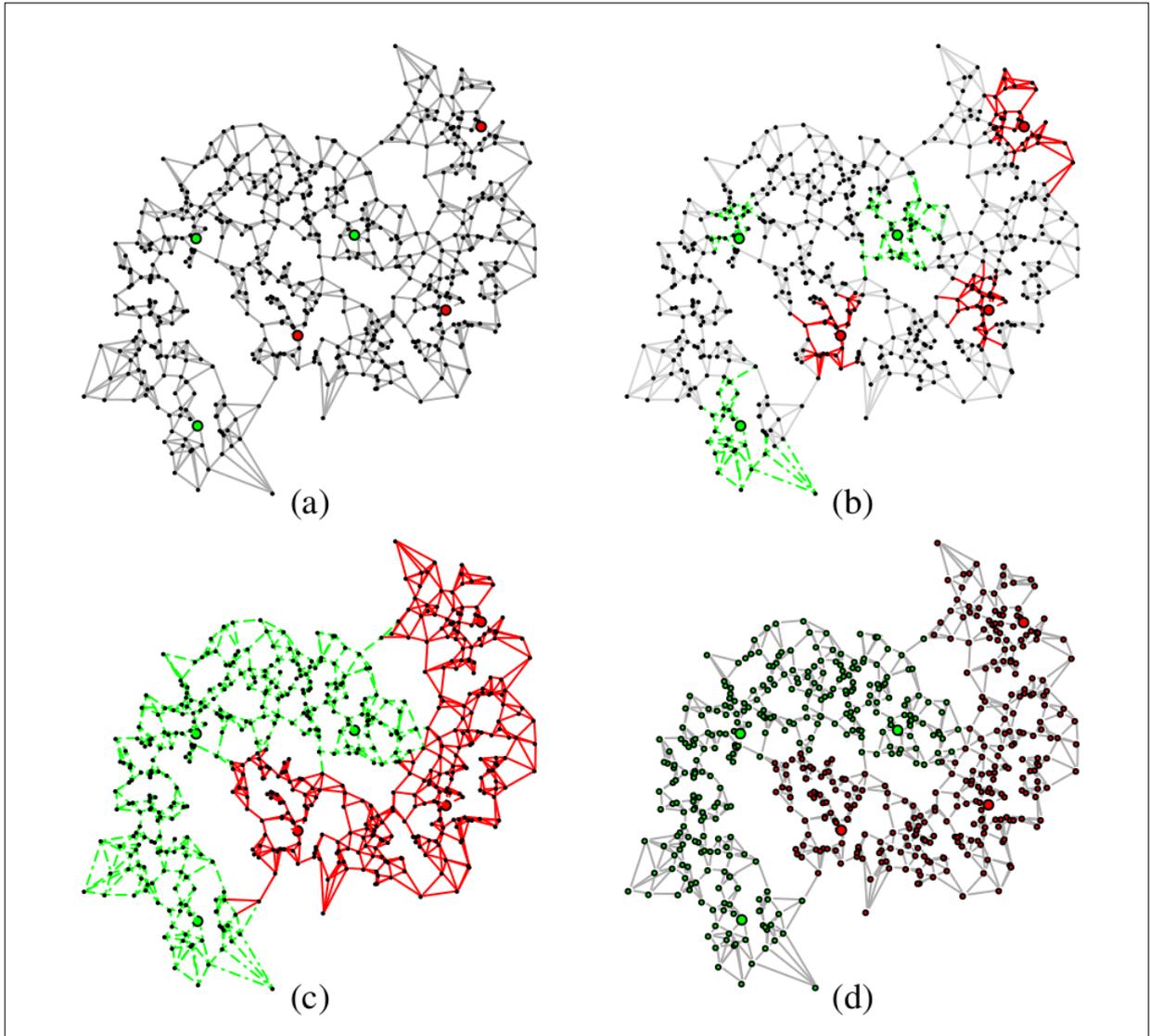
$$G^c(t) := (\mathcal{V}, \mathcal{E}^c(t)). \quad (2.13)$$

Por fim, as redes  $G^c(\tau)$  são usadas para classificação dos vértices. São atribuídos os rótulos  $y_j \in \{1, \dots, C\}$  para cada vértice não rotulado com a informação provida pela rede  $G^c$ . O rótulo  $y_j$  é atribuído baseado na densidade das arestas na sua vizinhança. Formalmente, escrito como:

$$y_j = \arg \max_{c \in \{1, \dots, C\}} |\mathcal{E}(\mathcal{N}_{c,j})|, \quad (2.14)$$

em que  $\mathcal{N}_{c,j}$  são as arestas da vizinhança do vértice  $j$  que foram dominadas por partículas da classe  $c$ . Na Figura 14, é ilustrada a dinâmica coletiva aplicada em uma rede complexa de formato similar a duas bananas. A ilustração está dividida em quatro etapas: (a) inicialização com 6 vértices rotulados, 3 para cada classe, com vértices pretos não rotulados; (b) 4 iterações do sistema, com as arestas coloridas sendo dominadas por partículas, enquanto as cinzas ainda não foram dominadas; (c) 20 iterações, todo o grafo foi percorrido; (d) a classificação dos vértices pela classe correspondente que obteve maior dominação.

Figura 14 – Execução em quatro fases da dinâmica coletiva LCU.



Fonte: Verri *et al.* (2018)

## 2.8 EGSIS

O algoritmo *Exploratory Graph-based Semi-supervised Image Segmentation* (EGSIS) é a técnica de segmentação de imagens proposta neste trabalho. Todas as seções anteriores da fundamentação teórica foram introduzidas para explicar o funcionamento dessa técnica que será descrita nesta seção.

O modelo EGSIS permite a execução de uma segmentação transdutiva de imagens na presença de uma rotulação parcial da imagem. Essa técnica possui etapas flexíveis para refinamento e otimização, quanto a método de extração de características, quantidade de superpixels, função de similaridade, entre outros.

Ao considerar as características do algoritmo LCU, explicadas na Seção 2.7,

esse algoritmo também pode ser usado para aprendizado semi-supervisionado multi-classes.

O ponto central desse método de segmentação é realizar uma transformação do domínio da imagem para uma rede complexa que seja compatível com a dinâmica coletiva LCU. Para isso, uma pré-segmentação é realizada usando superpixels através do algoritmo SLIC. Os rótulos iniciais da imagem são propagados aos superpixels baseado numa votação: a classe que tiver maior números de pixels rotulados é atribuída ao superpixel.

Ao obter os superpixels e os rótulos iniciais, é construída uma rede complexa onde os vértices representam os superpixels e as arestas a vizinhança com grau de similaridade entre os superpixels. Para calcular a similaridade de cada aresta, primeiro é executado para cada superpixel o método de extração de características selecionado, que nesse caso pode ser filtros de Gabor ou matrizes de co-ocorrência, então calcula-se a similaridade entre os superpixels com uma métrica de similaridade selecionada.

Nessa etapa, a rede complexa é preparada para execução do sistema dinâmico LCU a fim de propagar os rótulos no grafo. No final da execução do LCU, é reconstruída uma máscara de segmentação baseado nas rotulações dos vértices restantes em cada superpixel não rotulado. Por fim, obtém-se uma matriz de rótulos relacionando cada pixel ao rótulo de segmentação desejado.

No Algoritmo 2 são descritas as etapas do funcionamento geral do método.

---

**Algoritmo 2:** EGSIS

---

**Entrada:** Imagem, rotulações parciais

**Resultado:** Imagem segmentada

**início**

Inicializa os hiperparâmetros do modelo EGSIS;  
 Segmenta a imagem em  $n$  superpixels usando SLIC;  
 Realiza a construção da rede complexa baseado nos superpixels;  
 Atribui os superpixels rotulados baseado na anotação parcial;  
 Calcula o vetor de características para cada superpixel;  
 Calcula a similaridade entre superpixels da vizinhança;  
 Executa a dinâmica LCU sobre a rede complexa gerada;  
 Reconstrói a máscara de segmentação com base nos vértices rotulados;

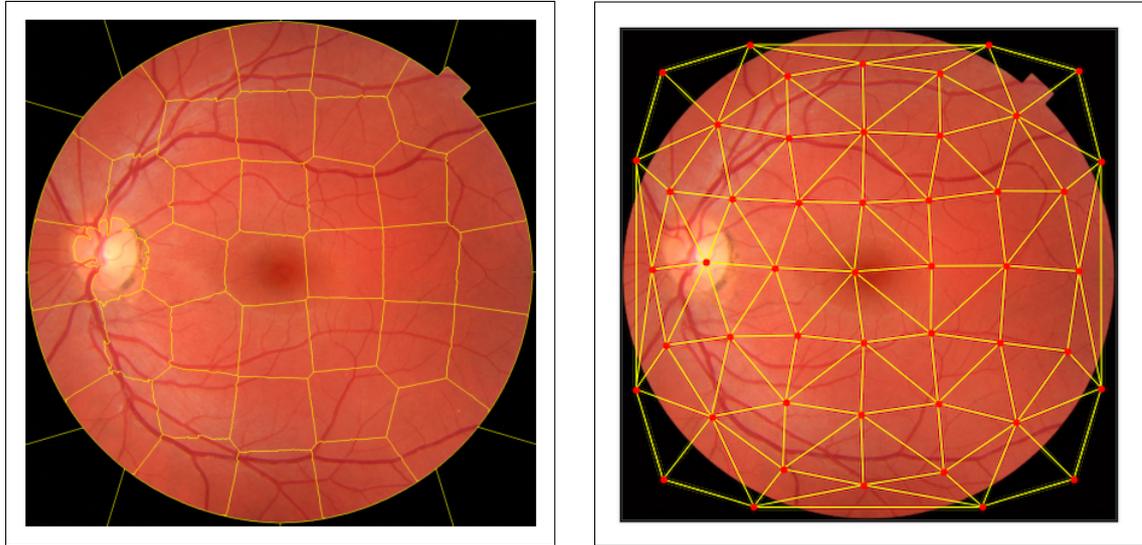
**fim**

---

A seguir, uma ilustração da execução do algoritmo em algumas etapas. Na

Figura 15, são observadas duas etapas sendo realizadas: na imagem (a) a segmentação por superpixels, na imagem (b) a rede complexa gerada através dos superpixels.

Figura 15 – Construção de superpixels e a rede complexa no modelo EGSIS.



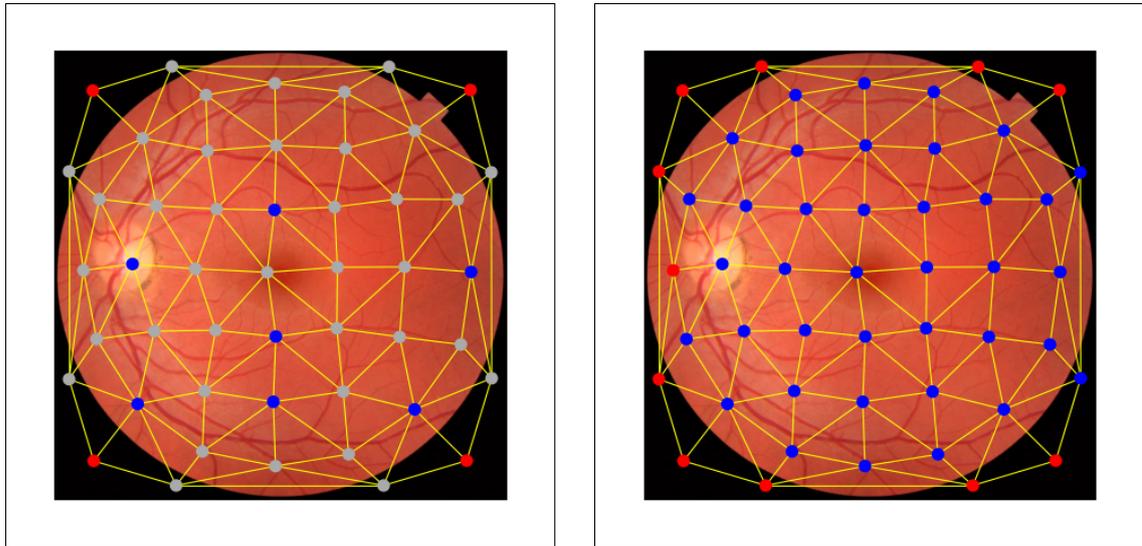
(a)

(b)

Fonte: Elaborado pelo autor baseado nas imagens disponíveis em Walt *et al.* (2014).

Na Figura 16, há alguns superpixels anotados em azul e vermelho, e uma segmentação em duas classes. Na outra imagem, é ilustrado o resultado após a continuação da execução. Perceba que boa parte da imagem foi segmentada corretamente, mas houve alguns erros visíveis.

Figura 16 – Etapa do método EGSIS para propagar rótulos iniciais através do sistema dinâmico LCU.



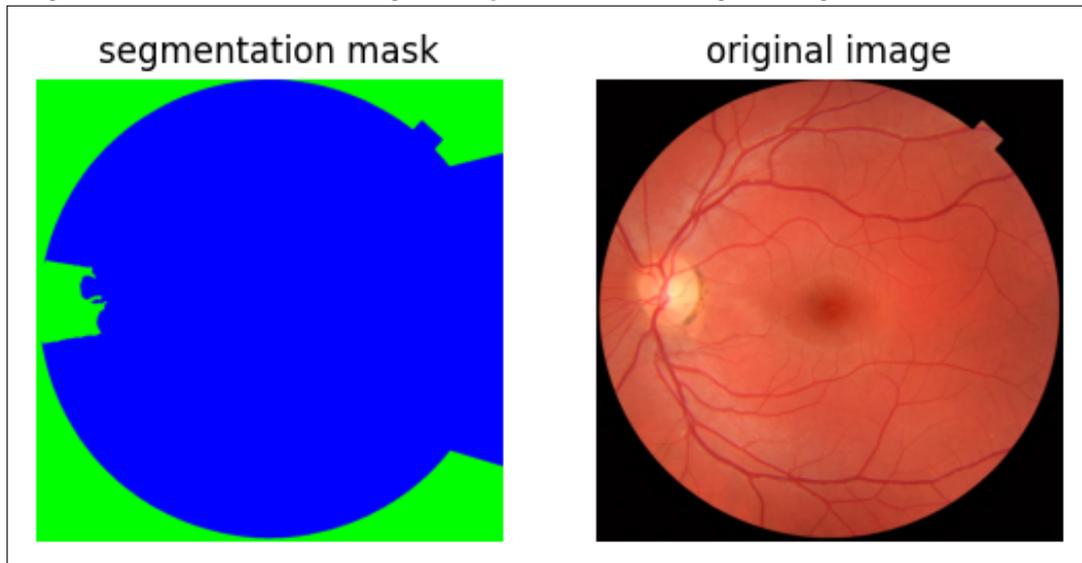
(a)

(b)

Fonte: Elaborado pelo autor baseado nas imagens disponíveis em Walt *et al.* (2014).

Por fim, obtém-se a seguinte máscara de segmentação ao atribuir os rótulos dos vértices para os superpixels iniciais, assim sendo possível reconstruir uma máscara de segmentação como ilustrado na Figura 17.

Figura 17 – Máscara de segmentação final e a imagem original.

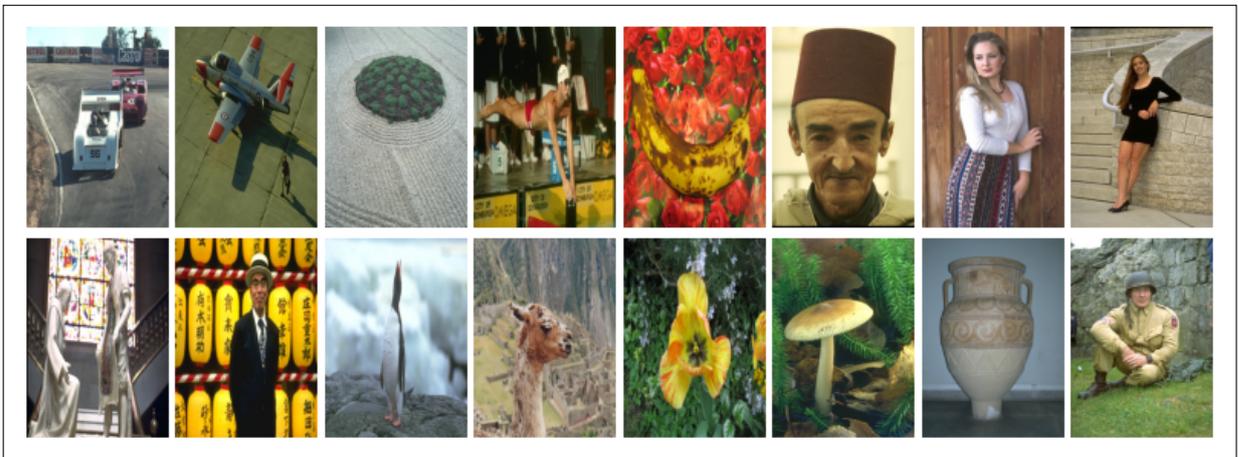


Fonte: Elaborado pelo autor baseado nas imagens disponíveis em Walt *et al.* (2014).

### 3 METODOLOGIA

Para avaliação deste trabalho, foi considerada a base de dados GrabCut (ROTHER *et al.*, 2004), que contém 50 imagens com segmentação binária e anotações parciais para execução de segmentações assistidas, assim como segmentação interativa. Nessa coleção de imagens, há variados tipos de objetos, como pessoas, carros, plantas, animais, entre outros tipos de imagem. Na Figura 18 é apresentada uma amostra de 16 imagens que serão usadas para realizar uma avaliação qualitativa do método EGSIS.

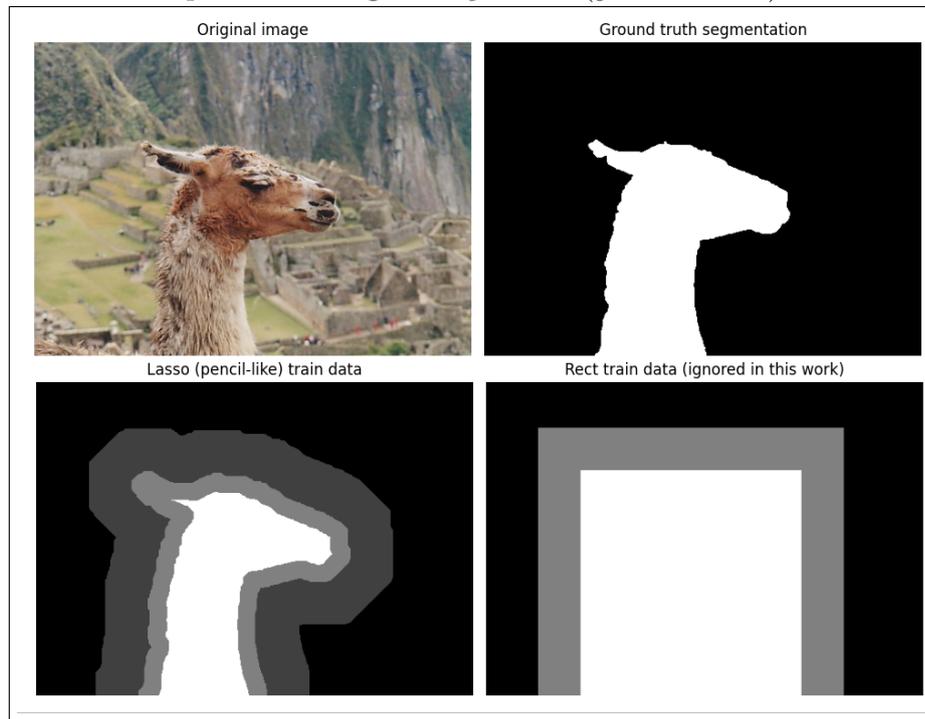
Figura 18 – Amostra de 16 imagens do *dataset* GrabCut.



Fonte: Elaborado pelo autor a partir de imagens do *dataset* GrabCut (ROTHER *et al.*, 2004).

As anotações parciais que serão consideradas neste trabalho são chamadas de *Lasso* e simulam um usuário realizando uma marcação de regiões do plano de fundo e primeiro plano para segmentação. A anotação possui um esquema de anotação para treinamento em escala de cinza e será descrita em sequência.

Figura 19 – Exemplo de imagem *dataset* GrabCut com anotação parcial e a segmentação real (*ground truth*).



Fonte: Elaborado pelo autor baseado em Rother *et al.* (2004)

Tabela 1 – Regras da anotação Lasso do GrabCut

Descrição	Nível de cinza
fundo	0
fundo usado para <b>treinamento</b>	64
região de inferência	128
primeiro plano usado para <b>treinamento</b>	255

Fonte: (ROTHER *et al.*, 2004)

Através da Figura 19, ao observar-se a primeira imagem na segunda linha, tem-se a anotação *Lasso*, na qual o significado dos níveis de cinza pode ser visualizado na Tabela 1. Dessa maneira, para treinamento<sup>1</sup> são usados dois níveis de cinza: 64 para o fundo e 255 para o primeiro plano.

O método de avaliação consiste em executar o algoritmo desenvolvido EGSIS para essa base de dados e extrair métricas de qualidade de segmentação para comparação com outros trabalhos, assim como estudo da variação de parâmetros do modelo e a percepção de impacto nas métricas de avaliação.

<sup>1</sup> EGSIS é um algoritmo transdutivo, não existe treinamento, esses dados são apenas anotações parciais que serão usadas na execução da inferência para obter a segmentação da imagem.

### 3.1 Métricas de avaliação

Para avaliação deste trabalho, foram selecionadas as seguintes métricas que são comumente utilizadas para avaliação de segmentação de imagens:

1. *Precision*;
2. *Recall*;
3. *F1 Score (Dice Coefficient)*;
4. IoU.

#### 3.1.1 *Precision*

Para avaliação da precisão da segmentação, tem-se a seguinte definição:

$$Precision = \frac{TP}{TP + FP}, \quad (3.1)$$

em que  $TP$  são os pixels classificados corretamente e  $FP$  os falsos-positivos. Essa métrica penaliza a presença de falsos-positivos. Quanto menos falsos-positivos, maior será a precisão.

#### 3.1.2 *Recall*

Para avaliação do *Recall* (também conhecido como Sensibilidade), tem-se a seguinte definição:

$$Recall = \frac{TP}{TP + FN}. \quad (3.2)$$

Nessa equação, similar à precisão, é introduzido  $FN$  no denominador, os falsos-negativos. Diferentemente da precisão, o *recall* penaliza a presença de falsos-negativos. Quanto menos falsos-negativos, maior será o *recall*.

#### 3.1.3 *F1 Score*

A métrica F1 Score, em segmentação de imagens também conhecida como *Dice Coefficient*, é uma soma harmônica entre *precision* e *recall*. Nessa situação, tem-se a definição:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \quad (3.3)$$

F1 score é uma métrica balanceada que, entre *precision* e *recall*, penaliza a que tiver pior avaliação.

### 3.1.4 IoU

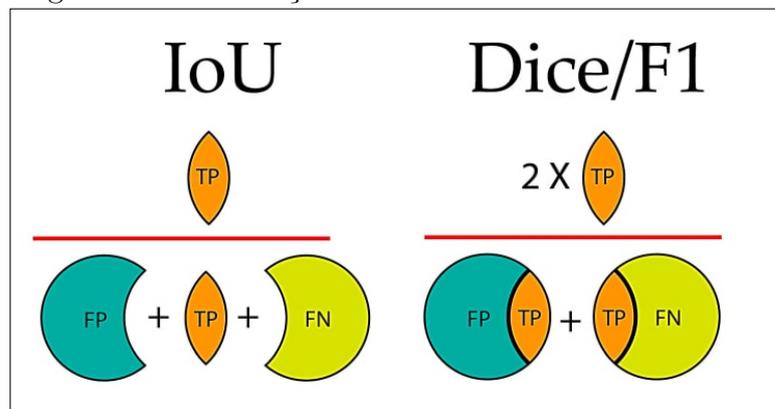
IoU (REZATOFIGHI *et al.*, 2019), que significa *Intersection over Union*, é uma métrica popularmente usada para medir a precisão da segmentação de um objeto em tarefas de visão computacional, como detecção de objetos e segmentação semântica.

A métrica IoU calcula a proporção da área de interseção entre a região estimada e a região real (*ground truth*) pela área da união dessas duas regiões. A fórmula para calcular o IoU é:

$$IoU = \frac{|A \cap B|}{|A \cup B|}. \quad (3.4)$$

Nessa equação,  $A$  e  $B$  são matrizes de rótulos da imagem que serão comparadas, sendo a matriz  $A$  os rótulos estimados e a matriz  $B$  os reais. O valor de IoU varia de 0 a 1, com o valor 1 indicando uma correspondência perfeita entre as regiões delimitadoras previstas e reais, e o valor 0 indicando que não há sobreposição (pior valor). Em geral, um IoU maior indica uma melhor precisão do modelo de segmentação. Para ilustração, tem-se a seguinte Figura 20:

Figura 20 – Ilustração das métricas IoU e F1 score.



Fonte: Maxwell *et al.* (2021)

## 4 RESULTADOS

Uma implementação do algoritmo EGSIS foi realizada como um projeto de código-aberto e está disponível na plataforma github<sup>1</sup>. Nessa seção de resultados, avalia-se o comportamento do método com variações no número de superpixels e método de extração de características. Como referência para comparação dos resultados, considera-se os métodos avaliados em Wang *et al.* (2023) no *dataset* GrabCut.

Uma avaliação qualitativa é demonstrada ao realizar a segmentação do método e disponibilizar a máscara de segmentação gerada para algumas imagens. Além disso, em formato de tabelas, é disponibilizada a avaliação quantitativa baseada nas métricas descritas na Seção 3.1.

Por fim, foi desenvolvida uma ferramenta simples de anotação para exploração dos resultados integrada ao software JupyterLab<sup>2</sup>, que é usada também para avaliação qualitativa e visualização do impacto da anotação parcial inicial no resultado final.

### 4.1 Experimentos com a variação da quantidade de superpixels

Durante a execução dos experimentos, e avaliação dos resultados, foi observado que o número de superpixels afeta diretamente a qualidade da segmentação gerada. Entre esses experimentos, foram realizadas 8 execuções com variações na técnica de extração de características usada e no número de superpixels, sendo o número de superpixels avaliados para 50, 100, 150 e 200.

Tabela 2 – Resultados dos experimentos ao variar o número de superpixels e o método de extração de características. Em negrito os melhores resultados.

Método	Recall	Precision	F1	IoU	Tempo (s)
<i>EGSIS</i> ( $s = 50, f = gabor$ )	0,7882	0,9225	0,8414	0,7412	9,22
<i>EGSIS</i> ( $s = 100, f = gabor$ )	0,8592	0,9505	0,8992	0,8232	14,78
<i>EGSIS</i> ( $s = 150, f = gabor$ )	0,8691	0,9550	0,9066	0,8354	18,44
<i>EGSIS</i> ( $s = 200, f = gabor$ )	0,8799	<b>0,9620</b>	0,9167	0,8511	21,06
<i>EGSIS</i> ( $s = 50, f = comatrix$ )	0,7882	0,9225	0,8414	0,7412	<b>4,94</b>
<i>EGSIS</i> ( $s = 100, f = comatrix$ )	0,8617	0,9510	0,9010	0,8259	11,26
<i>EGSIS</i> ( $s = 150, f = comatrix$ )	0,8777	0,9552	0,9125	0,8441	15,12
<i>EGSIS</i> ( $s = 200, f = comatrix$ )	<b>0,8877</b>	0,9611	<b>0,9212</b>	<b>0,8578</b>	19,08

Fonte: Elaborado pelo autor.

<sup>1</sup> Disponível em <<https://github.com/ryukinix/egsis/releases/tag/0.1.0>>

<sup>2</sup> Software web para desenvolvimento interativo e análise de dados, disponível em <<https://jupyter.org>>

Ao analisar a Tabela 2, é possível perceber que o aumento de superpixels impacta positivamente na qualidade da segmentação, independente do método de extração de características. A respeito do método de extração de características, os resultados de ambos os métodos são próximos. Por outro lado, com exceção da métrica de precisão, os melhores resultados vieram do modelo com 200 superpixels e extração de características via matriz de co-ocorrências. O aumento de superpixels também contribuiu para o aumento no tempo de execução do modelo. Não foi possível observar melhora significativa nas métricas de avaliação ao usar mais de 200 superpixels. Também é importante salientar que a melhora mais significativa ocorre ao aumentar de 50 para 100 superpixels. Para poucos superpixels, o tempo de execução é significativamente menor ao usar matriz de co-ocorrências em vez de filtros de Gabor.

Todos os experimentos deste trabalho foram executados em um notebook pessoal modelo *Thinkpad E480*, com processador *Intel i5-8250U @ 3.400GHz*, 16GB de memória *RAM* e sistema operacional *Ubuntu 22.04.3 LTS*.

## 4.2 Avaliação qualitativa

Nesta avaliação qualitativa, é comparada visualmente a qualidade de segmentação do método EGSIS com os melhores parâmetros encontrados na seção anterior, isto é, número de superpixels definido para 200 e método de extração de características via matriz de co-ocorrências.

Como referência, o artigo (WANG *et al.*, 2023) contém execuções comparativas de vários métodos na base de dados GrabCut. Neste artigo, é selecionada uma comparação com 16 imagens, nas quais 7 diferentes métodos de segmentação interativa foram usados.

Figura 21 – Comparação entre diferentes métodos de segmentação interativa no dataset GrabCut.



Fonte: Wang *et al.* (2023)

Na Figura 21, há em cada linha duas imagens a serem segmentadas. As máscaras de segmentação binária, em que a cor preta representa o plano de fundo e a cor branca o objeto de interesse, são os resultados das segmentações dos métodos de segmentação interativa. O nome desses 7 métodos, em ordem, são: GrabCut, LazySnapping, OneCut, Saliency Cut, Iterated Graph Cuts<sup>3</sup>, DenseCut e Deep GrabCut.

Nesse aspecto, para comparação com o modelo proposto EGSIS, para as mesmas imagens do dataset GrabCut, são obtidas as seguintes máscaras de segmentação representadas na Figura 22.

<sup>3</sup> Não há um nome formal pra esse método (AN; PUN, 2013), por ora ele será chamado assim.



### 4.3 Comparação quantitativa com o estado-da-arte da segmentação interativa

Nesta seção, é proposta uma avaliação de como o modelo EGSIS se compara ao estado-da-arte no problema de segmentação interativa.

No artigo de Wang *et al.* (2023), é observado um extenso estudo de revisão sobre os métodos de segmentação interativa baseados em grafos, tendo como base o algoritmo GrabCut e outros algoritmos que evoluíram nessa área de segmentação interativa de imagens. Os autores avaliam essas técnicas na base de dados GrabCut (ROTHER *et al.*, 2004) utilizando as mesmas métricas que foram selecionadas na metodologia de avaliação deste trabalho, com exceção do tempo de execução, que, por conta de diferentes tecnologias usadas para implementação das técnicas, não foi possível usar para fazer uma avaliação justa, por isso será deixada como um potencial trabalho futuro.

Esse mesmo artigo também avalia as várias técnicas em outro *dataset*, mas em virtude de a rotulação inicial usada não ter sido publicada, não foi possível replicar os experimentos para o modelo EGSIS nas mesmas condições. Por esse motivo, a avaliação de resultados será limitada a base de dados GrabCut.

A seguir, é possível visualizar tabelas de comparação com os algoritmos apresentados em Wang *et al.* (2023), todos algoritmos de segmentação interativa baseado em grafos.

Tabela 3 – Resultados comparativos entre o método EGSIS e métodos estado-da-arte para segmentação interativa baseado em grafos. Em negrito os melhores resultados, em vermelho os piores.

Método	Recall	Precision	F1	IoU
GrabCut	0,9668	0,9213	<b>0,9407</b>	<b>0,8927</b>
LazySnapping	<b>0,9681</b>	0,9104	0,9357	0,8842
OneCut	0,8585	<b>0,7926</b>	<b>0,7899</b>	<b>0,6974</b>
Saliency Cuts	<b>0,8371</b>	0,8892	0,8255	0,7458
Iterated Graph Cuts	0,9614	0,8878	0,9212	0,8597
DenseCut	0,8427	0,9418	0,8561	0,7927
Deep GrabCut	0,8854	0,8774	0,8701	0,7849
<i>EGSIS</i> ( $s = 200$ , $f = gabor$ )	0,8799	<b>0,9620</b>	0,9167	0,8511
<i>EGSIS</i> ( $s = 200$ , $f = comatrix$ )	0,8877	0,9611	0,9212	0,8578

Fonte: Baseado nos resultados encontrados no artigo de Wang *et al.* (2023)

Tabela 4 – Resultados do modelo EGSIS com desvio-padrão.

<b>Métrica</b>	<i>EGSIS</i> ( $s = 200, f = gabor$ )	<i>EGSIS</i> ( $s = 200, f = comatrix$ )
<b>Recall</b>	0,8799 $\pm$ 0,0907	0,8877 $\pm$ 0,0773
<b>Precision</b>	0,9620 $\pm$ 0,0314	0,9611 $\pm$ 0,0347
<b>F1</b>	0,9167 $\pm$ 0,0581	0,9212 $\pm$ 0,0506
<b>IoU</b>	0,8511 $\pm$ 0,0942	0,8578 $\pm$ 0,0838

Fonte: Elaborado pelo autor.

Na Tabela 3, é possível observar que o método *EGSIS*( $s = 200, f = gabor$ ) atinge o maior valor médio de precisão entre todos os métodos comparados, com um valor de 0,9620. Seguindo o método EGSIS, o segundo colocado na métrica de precisão foi o método DenseCut, com um valor de precisão de 0,9418, que é próximo, mas ainda significativamente menor. É importante ressaltar que uma alta precisão implica uma baixa incidência de pixels falsos-positivos.

Na Tabela 4, é possível visualizar os resultados da média e do desvio-padrão de cada métrica de avaliação. Ao fazer uma análise incluindo o desvio-padrão, não se pode concluir estatisticamente qual melhor método de extração de características, seja a matriz de co-ocorrências ou os filtros de Gabor. Em todas as métricas há uma intersecção dos valores médios quando se considera o desvio-padrão.

Em nenhuma das métricas apresentadas na comparação o EGSIS se classifica como o pior método. No entanto, observa-se um valor menor de *recall* em comparação à precisão, o que impacta diretamente no valor das métricas *F1* e *IoU*. Neste cenário, conclui-se que o modelo EGSIS tem uma taxa maior de falsos-negativos do que de falsos-positivos.

Embora seja necessária uma investigação mais aprofundada para determinar a causa dessa maior taxa de falsos-negativos, é importante notar que o conjunto de dados GrabCut possui um desequilíbrio na quantidade de pixels entre as classes, sendo o plano de fundo a classe dominante. No algoritmo LCU, um dos fatores importantes para classificar um vértice que representa um superpixel é a extração de características, que por meio de uma função de similaridade é usada para o cálculo do peso das arestas da rede complexa. No entanto, esse é apenas um dos fatores do sistema dinâmico, pois também há a geração de partículas e a transição entre os vértices. Geralmente, o objeto é cercado pelo plano de fundo. Isso pode levar a uma situação na qual a rede complexa maior, localizada fora do objeto, gera partículas da classe do plano de fundo em quantidade suficiente para dominar

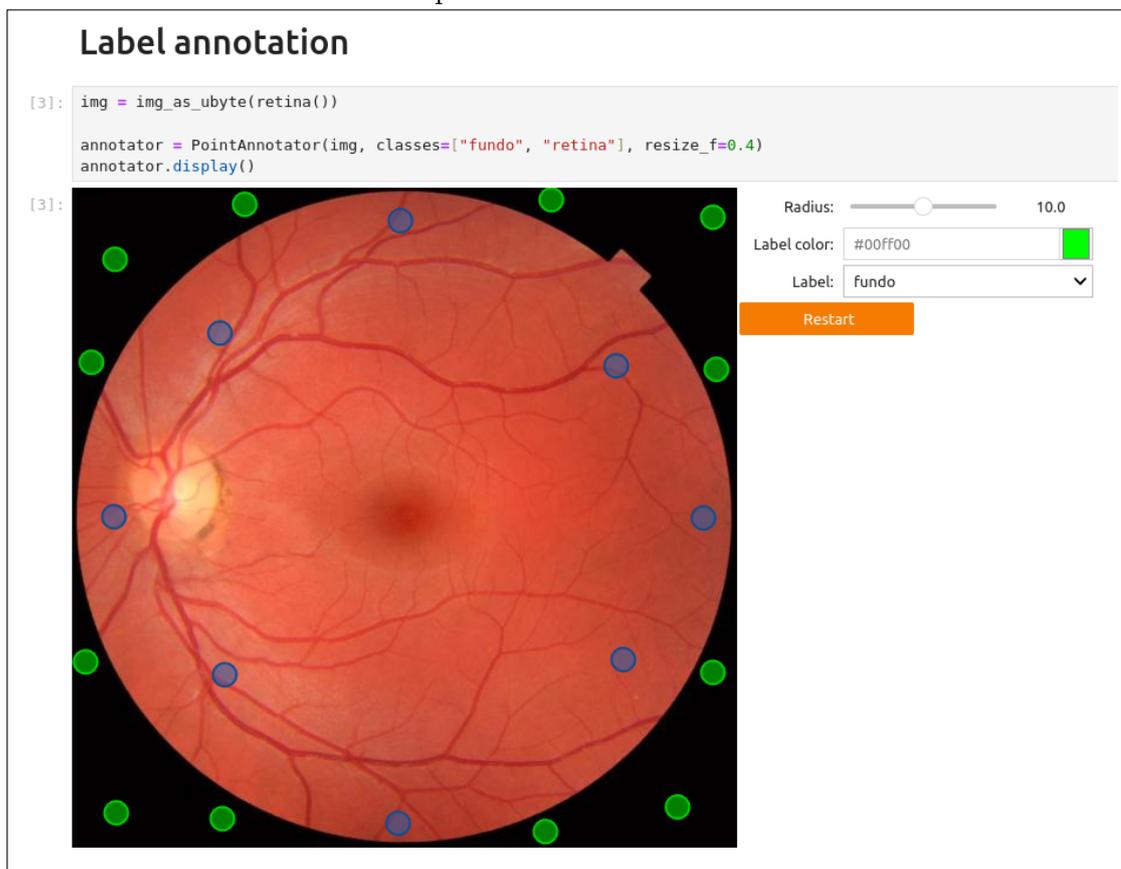
os vértices na região do objeto. Esse comportamento implicaria um aumento na incidência de falsos-negativos.

Se isso for confirmado, uma possível contramedida seria modificar o sistema dinâmico para dar maior peso às similaridades entre superpixels, de tal maneira que haja um aumento da probabilidade de sobrevivência das partículas entre superpixels de mesma classe. Além disso, seria útil explorar novas técnicas de extração de características que gerem valores de similaridade mais discriminantes.

#### 4.4 Ferramenta de anotação integrada com método EGSIS

Para a exploração e avaliação dos resultados a nível individual de imagem, desenvolveu-se uma ferramenta de anotação parcial de imagens para testar o método EGSIS. Esta ferramenta é utilizada no cenário de segmentação interativa e seu uso é descrito nesta seção.

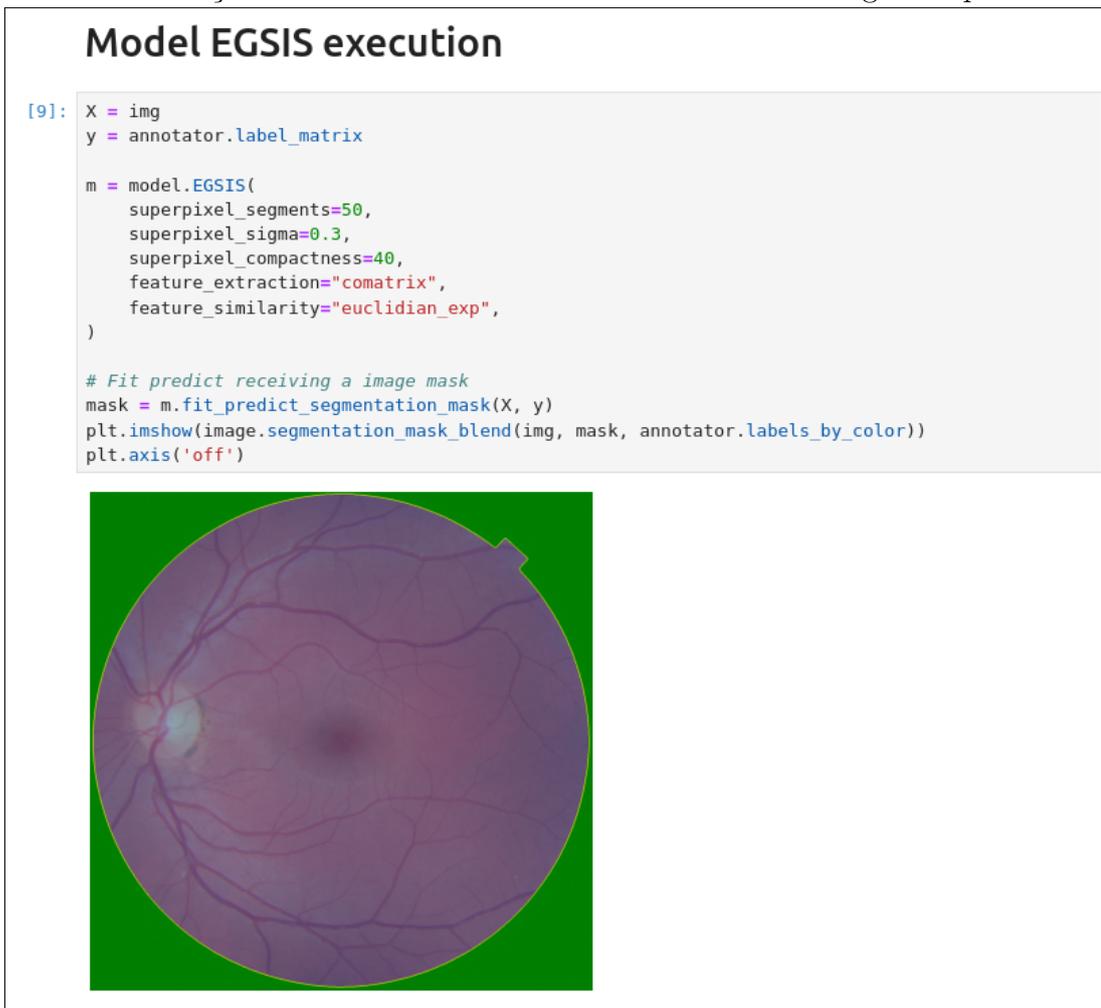
Figura 23 – Ferramenta para anotação parcial de imagem, com pontos circulares e classes distintas por cor.



Fonte: Elaborado pelo autor baseado nas imagens disponíveis em Walt *et al.* (2014).

Na Figura 23 é ilustrada a ferramenta desenvolvida. Durante a inicialização do anotador, é possível declarar o nome das classes a serem anotadas e selecionar a cor das anotações — neste caso, verde para o plano de fundo e azul para a retina. A cada clique do mouse, um ponto circular com raio variável pode ser anotado, cujo tamanho pode ser definido pela ferramenta através de um botão deslizante. Por fim, obtém-se uma instância do anotador que contém as anotações e que pode ser usada para gerar uma matriz de rótulos por pixel.

Figura 24 – Execução do método EGSIS com a matriz de rótulos gerada pelo anotador.



Fonte: Elaborado pelo autor baseado nas imagens disponíveis em Walt *et al.* (2014).

Na Figura 24, a execução do algoritmo EGSIS é ilustrada utilizando os dados de anotação como matriz de rótulos  $y$ , que contém as anotações parciais. Ao final da execução, a máscara de segmentação colorida é exibida sobreposta à imagem original. Nessa situação, uma segmentação perfeita foi alcançada após corrigir os erros do modelo na ferramenta de anotação com novos pontos.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, foi possível conceber um novo algoritmo chamado de EGSIS que pode ser enquadrado na resolução de problemas de segmentação interativa de imagens. Ao realizar a pesquisa sobre esse tema, foi observado que a linha de pesquisa sobre algoritmos transdutivos de segmentação de imagens ainda está evoluindo, mas com menos volume que as técnicas indutivas e, por esse motivo, este trabalho pode se posicionar como uma referência para construção de mais trabalhos envolvendo técnicas similares.

Na seção de resultados, o algoritmo EGSIS obteve uma performance próxima daquelas obtidas pelos métodos estado-da-arte baseados em grafos para segmentação interativa de imagens. Durante a exploração dos resultados, foi observado que a qualidade da segmentação final depende crucialmente da qualidade da pré-segmentação com superpixels.

Durante os experimentos, foi observado também que os melhores resultados foram alcançados com um maior número de superpixels e que o aumento do número de superpixels aumentou também o tempo de execução do algoritmo.

O algoritmo possui várias partes móveis e trocáveis, o que permite no futuro a evolução de técnicas para melhorar ainda mais a qualidade de segmentação final. Os métodos de superpixel, extração de características, construção da rede complexa e dinâmica coletiva são partes móveis que podem ser testadas com técnicas mais avançadas.

### 5.1 Limitações e desafios

Um ponto de dificuldade em toda a pesquisa foi encontrar trabalhos de referência nessas condições com a palavra-chave semi-supervisionado, uma vez que as buscas realizadas direcionavam para trabalhos nos quais os algoritmos eram indutivos. Após trocar para segmentação transdutiva e segmentação interativa (mais recente), foi possível encontrar trabalhos que poderiam se enquadrar melhor com a proposta deste trabalho.

A falta de *datasets* para segmentação interativa com a inclusão da anotação parcial inicial foi um dos maiores desafios deste trabalho, o que limitou a capacidade de exploração e comparação com outros datasets de segmentação de imagens, assim como outras técnicas propostas em outros artigos. Por conta dessa ausência, o trabalho se limitou ao dataset GrabCut.

A segmentação interativa é um tipo de problema que, por natureza, é uma

segmentação binária. O algoritmo EGSIS permite a segmentação interativa, mas não é limitado a esta. O desenvolvimento da técnica considerou também segmentação multi-classes, mas, por conta da dificuldade de encontrar *datasets* públicos que contivessem as rotulações parciais, se tornou uma limitação da pesquisa.

A exploração na linha de pesquisa sobre algoritmos de segmentação transdutiva multi-classes ainda possui espaço para crescer. A criação de um dataset novo para *benchmark* que tivesse essas anotações do usuário, como existe no GrabCut, poderia ser uma contribuição relevante para essa linha de pesquisa.

## 5.2 Trabalhos futuros

Como mencionado anteriormente, o algoritmo EGSIS possui uma flexibilidade na troca de várias partes que o compõem. Como potenciais trabalhos futuros, uma das principais oportunidades de pesquisa seria testar técnicas estado-da-arte de extração de características de imagens. Neste trabalho, a extração de características foi limitada a matriz de co-ocorrência e filtros de Gabor, que não são as melhores técnicas atualmente, apesar de já terem sido suficientes para prover bons resultados.

A segmentação interativa evoluiu tanto para melhorar a qualidade de segmentação quanto para reduzir ao mínimo a necessidade de rotulações iniciais providas pelo usuário. Em publicações mais recentes sobre segmentação interativa, novos estudos buscam criar técnicas que minimizam o número de cliques de anotação que o usuário precisa fazer para alcançar segmentações de boa qualidade. Um trabalho futuro seria expandir as atuais métricas para simular a quantidade mínima de anotação necessária para alcançar um IoU com um determinado valor mínimo, geralmente entre 0,85 e 0,95.

## REFERÊNCIAS

- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. **SLIC superpixels**. [S.l.], 2010. Disponível em: <<https://www.epfl.ch/labs/ivrl/research/slic-superpixels/>>. Acesso em: 25 nov. 2023.
- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. Slic superpixels compared to state-of-the-art superpixel methods. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Institute of Electrical and Electronics Engineers, Los Alamitos, v. 34, n. 11, p. 8. 2274–2282, 2012. Disponível em: <<http://infoscience.epfl.ch/record/177415>>.
- AHMED, J. **Instance Segmentation with Mask R-CNN and Tensor-Flow on Onepanel**. 2019. Disponível em: <<https://medium.com/onepanel/instance-segmentation-with-mask-r-cnn-and-tensorflow-on-onepanel-6a072a4273dd>>. Acesso em: 27 mar. 2021.
- AN, N.-Y.; PUN, C.-M. Iterated graph cut integrating texture characterization for interactive image segmentation. In: IEEE. **2013 10th International Conference Computer Graphics, Imaging and Visualization**. [S.l.], 2013. p. 79–83.
- CHEN, X.; ZHAO, Z.; ZHANG, Y.; DUAN, M.; QI, D.; ZHAO, H. Focalclick: Towards practical interactive image segmentation. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2022. p. 1300–1309.
- COSTA, L. da F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: A survey of measurements. **Advances in Physics**, Informa UK Limited, v. 56, n. 1, p. 167–242, jan 2007. Disponível em: <<https://doi.org/10.1080%2F00018730601170527>>.
- DAUGMAN, J. G. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. **IEEE Transactions on acoustics, speech, and signal processing**, IEEE, v. 36, n. 7, p. 1169–1179, 1988.
- DUCHENNE, O.; AUDIBERT, J.-Y.; KERIVEN, R.; PONCE, J.; SÉGONNE, F. Segmentation by transduction. In: IEEE. **2008 IEEE conference on computer vision and pattern recognition**. [S.l.], 2008. p. 1–8.
- FENG, D.; HARAKEH, A.; WASLANDER, S. L.; DIETMAYER, K. A review and comparative study on probabilistic object detection in autonomous driving. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 23, n. 8, p. 9961–9980, 2021.
- HARALICK, R. M. Statistical and structural approaches to texture. **Proceedings of the IEEE**, IEEE, v. 67, n. 5, p. 786–804, 1979.
- HARALICK, R. M.; SHAPIRO, L. G. Image segmentation techniques. **Computer Vision, Graphics, and Image Processing**, v. 29, n. 1, p. 100–132, 1985. ISSN 0734-189X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0734189X85901537>>.
- IBRAHIM, A.; EL-KENAWY, E.-S. Image segmentation methods based on superpixel techniques: A survey. v. 1, p. 1–10, 10 2020.

KÄMÄRÄINEN, J.-K. Gabor features in image analysis. In: **International Conference on Image Processing Theory Tools and Applications**. [s.n.], 2012. Disponível em: <<https://api.semanticscholar.org/CorpusID:14909537>>.

LIMA, G. V. L. de; CASTILHO, T. R.; BUGATTI, P. H.; SAITO, P. T. M.; LOPES, F. M. A complex network-based approach to the analysis and classification of images. In: PARDO, A.; KITTLER, J. (Ed.). **Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications**. Cham: Springer International Publishing, 2015. v. 9423, p. 322–330. ISBN 978-3-319-25751-8.

LIRA, F. **Aplicação de Agrupamento Semissupervisionado para Segmentação de Imagens Coloridas**. Universidade Federal do Pernambuco (UFPE), 2018. Disponível em: <<http://www.bcc.ufrpe.br/sites/www3.bcc.ufrpe.br/files/Francisco%20Lira.pdf>>.

LUO, X.; CHEN, J.; SONG, T.; WANG, G. Semi-supervised medical image segmentation through dual-task consistency. In: **Proceedings of the AAAI conference on artificial intelligence**. [S.l.: s.n.], 2021. v. 35, n. 10, p. 8801–8809.

MAXWELL, A. E.; WARNER, T. A.; GUILLÉN, L. A. Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—part 1: Literature review. **Remote Sensing**, v. 13, n. 13, 2021. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/13/13/2450>>.

RAMADAN, H.; LACHQAR, C.; TAIRI, H. A survey of recent interactive image segmentation methods. **Computational visual media**, Springer, v. 6, p. 355–384, 2020.

REZATOFIGHI, H.; TSOI, N.; GWAK, J.; SADEGHIAN, A.; REID, I.; SAVARESE, S. Generalized intersection over union: A metric and a loss for bounding box regression. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition**. [S.l.: s.n.], 2019. p. 658–666.

RIBAS, L. C.; SÁ JÚNIOR, J. J. de M.; SCABINI, L. F.; BRUNO, O. M. Fusion of complex networks and randomized neural networks for texture analysis. **Pattern Recognition**, v. 103, p. 107189, 2020. ISSN 0031-3203. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0031320319304893>>.

ROTHER, C.; KOLMOGOROV, V.; BLAKE, A. Grabcut: interactive foreground extraction using iterated graph cuts. **ACM transactions on graphics (TOG)**, ACM New York, NY, USA, v. 23, n. 3, p. 309–314, 2004.

SAH, S. Machine learning: A review of learning types. **Preprints**, Preprints, July 2020. Disponível em: <<https://doi.org/10.20944/preprints202007.0230.v1>>.

SHAN, Y.; MA, Y.; LIAO, Y.; HUANG, H.; WANG, B. Interactive image segmentation based on multi-layer random forest classifiers. **Multimedia Tools and Applications**, Springer, v. 82, n. 15, p. 22469–22495, 2023.

SINGH, H. Image processing using machine learning. In: \_\_\_\_\_. **Practical Machine Learning and Image Processing: For Facial Recognition, Object Detection, and Pattern Recognition Using Python**. Berkeley, CA: Apress, 2019. p. 89–132. ISBN 978-1-4842-4149-3. Disponível em: <[https://doi.org/10.1007/978-1-4842-4149-3\\_5](https://doi.org/10.1007/978-1-4842-4149-3_5)>.

VAPNIK, V. Transductive inference and semi-supervised learning. In: CHAPELLE, O.; SCHÖLKOPF, B.; ZIEN, A. *et al.* (Ed.). **Semi-supervised Learning**. MIT: MIT press Cambridge, 2006. cap. 24, p. 453–472.

VAPNIK, V. N. **The nature of statistical learning theory**. [S.l.]: Springer-Verlag New York, Inc., 1995. 169 p. ISBN 0-387-94559-8.

VERRI, F. A. N.; URIO, P. R.; ZHAO, L. Network unfolding map by vertex-edge dynamics modeling. **IEEE Transactions on Neural Networks and Learning Systems**, Institute of Electrical and Electronics Engineers (IEEE), v. 29, n. 2, p. 405–418, Feb 2018. ISSN 2162-2388. Disponível em: <<http://dx.doi.org/10.1109/TNNLS.2016.2626341>>.

WALT, S. van der; SCHÖNBERGER, J. L.; Nunez-Iglesias, J.; BOULOGNE, F.; WARNER, J. D.; YAGER, N.; GOUILLART, E.; YU, T.; CONTRIBUTORS the scikit-image. scikit-image: image processing in Python. **PeerJ**, v. 2, p. e453, 6 2014. ISSN 2167-8359. Disponível em: <<https://doi.org/10.7717/peerj.453>>.

WANG, M.; LIU, X.; GAO, Y.; MA, X.; SOOMRO, N. Q. Superpixel segmentation: A benchmark. **Signal Processing: Image Communication**, v. 56, p. 28–39, 2017. ISSN 0923-5965. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0923596517300735>>.

WANG, Z.; LV, Y.; WU, R.; ZHANG, Y. Review of grabcut in image processing. **Mathematics**, MDPI, v. 11, n. 8, p. 1965, 2023.

ZHOU, B.; CHEN, L.; WANG, Z. Interactive deep editing framework for medical image segmentation. In: SPRINGER. **Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22**. [S.l.], 2019. p. 329–337.