

Automatização de oráculos de teste para imagens médicas de modelos tridimensionais

Misael Costa Júnior

Orientador: Prof. Dr. Márcio Delamaro

Universidade de São Paulo - USP

Instituto de Ciências Matemáticas e de Computação - ICMC

misaeljr@usp.br

19 de fevereiro de 2018



Agenda

- Introdução
- Oráculos de teste baseados em características
- Fluxo de atividades
- Conclusões
- Publicações

Introdução

Contexto do trabalho

- Engenharia de Software: produzir software de alta qualidade;
- Verificação, Validação e Teste (VV&T);
- Teste manual vs Teste automatizado;
- Oráculos de teste são fundamentais em cenários de testes automatizados ([OLIVEIRA et al., 2014a](#)).

Oráculos de teste

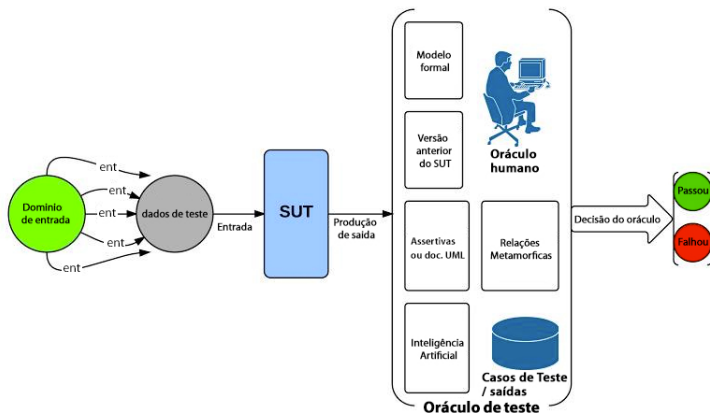


Figura: Fluxo genérico de teste com oráculos (Adaptado de OLIVEIRA (2017)).

Oráculos de teste: Problema do oráculo

- Não há oráculos de teste que sempre apoiam a decisão correta (WEYUKER, 1982; BARR et al., 2015);
- Ocorrência de falsos positivos e falsos negativos;
- O problema de oráculo se agrava em saídas complexas.



Figura: Exemplo de formatos de saídas que dificultam a automatização de oráculos.

Motivação – Imagens sintéticas 3D de vasos sanguíneos

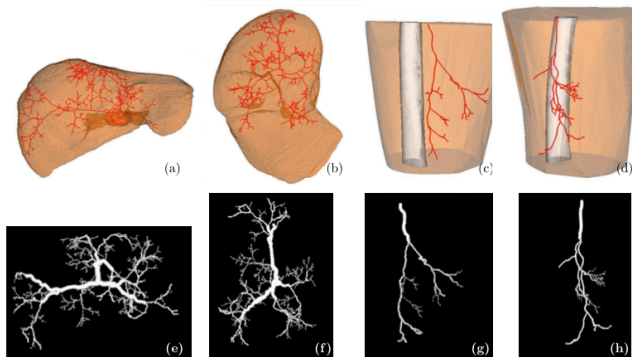


Figura: Vasos sintéticos 3D – antes e depois da segmentação (Adaptado de [GALARRETA-VALVERDE et al. \(2013\)](#)).

Objetivo

Fornecer alternativas para uma avaliação **criteriosa, sistemática e produtiva** da qualidade de sistemas com saídas que processem ou gerem imagens sintéticas tridimensionais de vasos sanguíneos, por meio de **oráculos automatizados**.

Objetivos específicos

- Definir o modo como extratores de características podem fornecer informações sobre a correção de sistemas que processem ou gerem imagens sintéticas 3D de vasos sanguíneos;
- Documentar e generalizar características e particularidades de sistemas que processem ou gerem imagens sintéticas 3D de vasos sanguíneos;
- Generalizar um *framework* para teste de sistemas de saídas complexas;
- Estabelecer um ambiente para que novos sistemas 3D sejam inseridos no contexto do *framework*.

Oráculos de teste baseados em características

Conceitos básicos

- Exploram características extraídas de uma saída confiável indicada pelo testador como referência;
- Utiliza conceitos de Recuperação de Imagem Baseada em Conteúdo (do inglês, *Content-Based Image Retrieval* - CBIR);
- São compostos de dois elementos básicos: (i) extração de características e (ii) comparações de objetos;
- Para projetar a abordagem é necessário que sejam explicitados os seguintes elementos: (i) o grupo de características relevantes; (ii) a função de similaridade para medir a distância entre as saídas SUT e as referências predefinidas; e o (iii) limiar de proximidade.

Um passo a passo

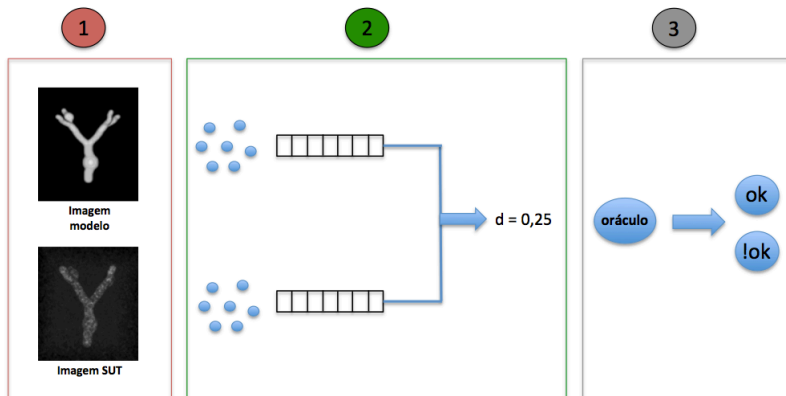


Figura: Abordagem oráculos de teste baseado em características – um passo a passo.

Framework O-Flm/CO - *Oracle for Images*

- Apoia testes em sistemas com saídas gráficas ou complexas;
- Emprega conceitos de CBIR;
- Oráculos gráficos (do inglês, *Graphical Oracle* - Gr-O) ([DELAMARO et al., 2013](#)).

Framework O-Flm/CO - *Oracle for Images*

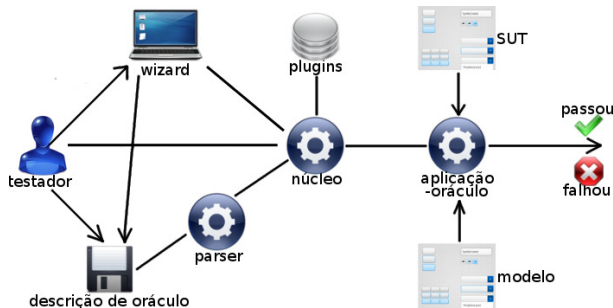


Figura: Estrutura genérica do O-Flm/CO (Adaptado de OLIVEIRA et al. (2014)).

Framework O-Flm/CO - *Oracle for Images*

- Extratores de características (*IExtractor*);
- Funções de similaridade (*ISimilarity*).

similarity Euclidean

extractor MyExtractor { color = "red" alpha = 78 rectangle = [100 100 30 40] }

extractor OurExtractor { rectangle = [0 0 128 64] scale = 1.33 }

precision = 0.46

Figura: Exemplo de um oráculo gráfico para o *framework* O-Flm/CO (Adaptado de [OLIVEIRA \(2012\)](#)).

Fluxo de atividades

Fluxo de atividades

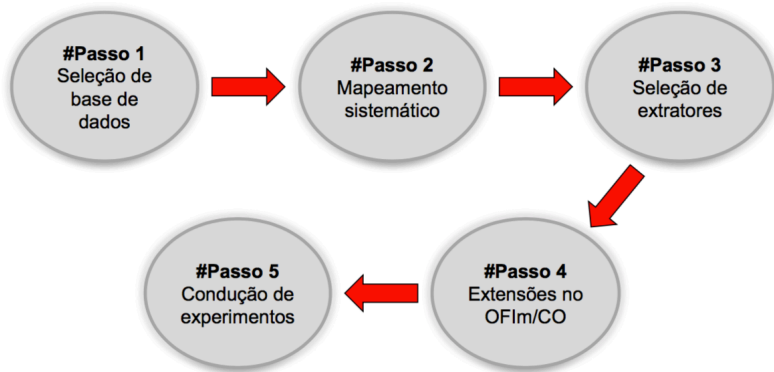


Figura: Fluxo de atividades do mestrado.

Passo 1: Seleção de base de dados



Figura: Fluxo de atividades do mestrado.

Imagens sintéticas 3D de vasos sanguíneos

- Muito úteis para exames de angio-RM ou por angio-TC;
- Sistemas que não têm sua qualidade mensurada devido à falta de estratégias de teste automatizado ([GALARRETA-VALVERDE et al., 2013](#));
- A atividade para modelagem de objetos tridimensionais (3D) em sistemas computacionais é complexa;
- Apesar da sua importância, exames de angiografia devem ser evitados:
 - Ataques cardíacos;
 - Derrames no paciente.

Exemplo de imagens

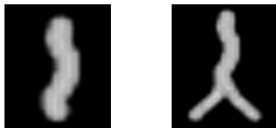
aneurisma



não-regular



normal



estenose



Figura: Categorias de imagens sintéticas de vasos sanguíneos.

Passo 2: Mapeamento sistemático (MS)



Figura: Fluxo de atividades do mestrado.

Objetivos

- Identificar métodos e abordagens de teste de software e análise de qualidade em sistemas biomédicos;
- Definir uma taxonomia de estratégias de teste de software e análise de qualidade em sistemas biomédicos;
- Obter uma visão geral dos procedimentos experimentais mais adequados para avaliação de sistemas biomédicos;
- Relatar as principais limitações e dificuldades na validação de sistemas biomédicos.

Processo de seleção

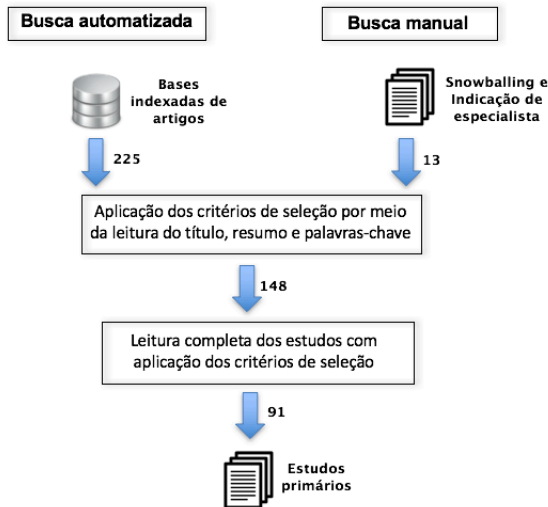
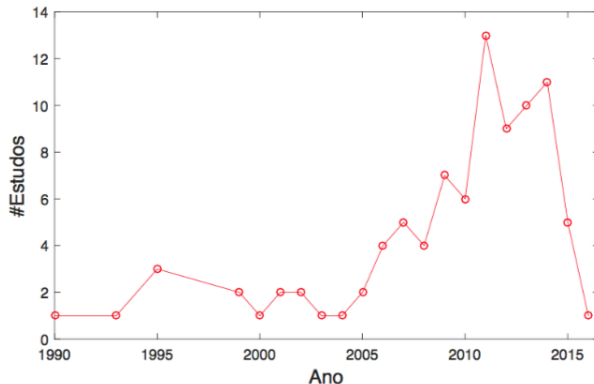


Figura: Processo de seleção (KITCHENHAM, 2004).

Seleção de estudos primários



Intervalo ano	#
1990-2000	8
2001-2005	8
2006-2010	26
2011-2016	49
Total	91

Figura: Análise de publicações por ano.

Sistemas que processam imagens médicas

- 23% (21/91) dos estudos apresentam abordagens para sistemas que geram ou processam imagens médicas;
- 85% (18/21) desses estudos discutem a complexidade do processo de avaliação nesse domínio de sistema;
- [FILHO et al. \(2014\)](#) e [GIBSON et al. \(2001\)](#) avaliam a qualidade de imagens médicas por meio de abordagens automatizadas, demonstrando que tais abordagens são mais eficientes em comparação ao uso de métodos manuais;
- Automatização de oráculos de teste para sistemas com saídas complexas ([OLIVEIRA et al., 2014b](#); [DELAMARO et al., 2013](#)).

Passo 3: Seleção e implementação de extratores



Figura: Fluxo de atividades do mestrado.

Processo de seleção e implementação

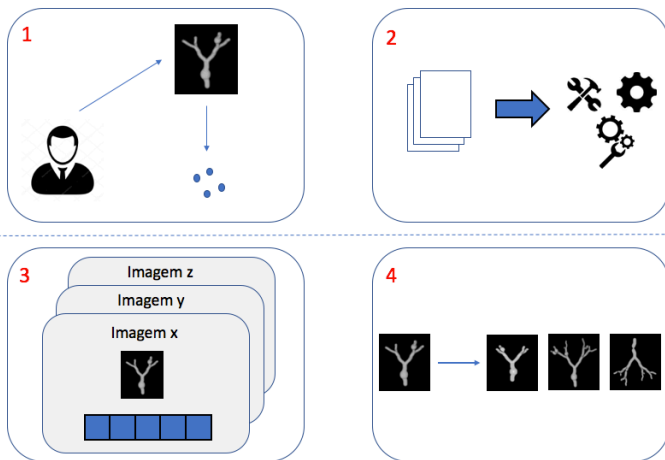
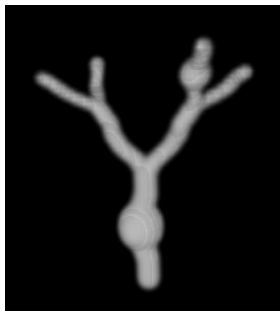


Figura: Processo de seleção e implementação dos extratores.

Exemplo



Características extraídas

Dimensões da imagem: 120 120 120

Extrator de densidade

Densidade: $1,03 \times 10^{-4}$

Extrator de bifurcação

Números de bifurcações: 3

Extrator de pontos terminais

Número de pontos terminais: 5

Extrator de segmentos

Número de segmentos: 7

Extrator de tamanho médio de segmentos

Tamanho médio de segmentos: 17,28

Figura: Exemplo de características obtidas por meio dos extratores desenvolvidos.

Passo 4: Extensões no OFIm/CO

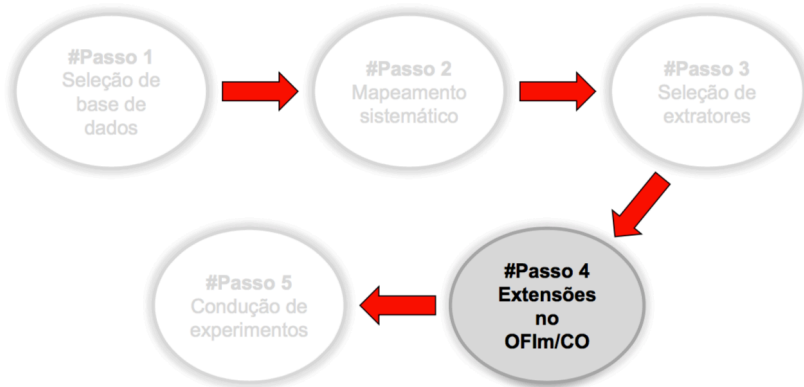


Figura: Fluxo de atividades do mestrado.

Extensões no *framework* OFIm/CO

- O *framework* OFIm/CO oferece uma interface Java completa para sistemas com saídas complexas chamada OFImPlugin;
- No contexto do presente trabalho, foi mantida a interface original, realizando uma pequena alteração no parâmetro do método **computeValue**;
- Duas interfaces Java para adaptar *plug-ins* do O-FIm/CO: (i) *InterfaceImageExtractor*; e (ii) *InterfaceSimilarityFunction*;
- Um catálogo de *plugins* foi desenvolvido e disponibilizado para imagens médicas 3D no formato NIfTI.

Extensões no *framework* OFIm/CO

Código-fonte 2 – Classe abstrata para *plug-ins* do O-Flm/CO

```
1: package br.oracle.pluginInterfaces;
2: import java.io.IOException;
3: public abstract class OFImPlugin {
4:     public abstract String getName();
5:     public abstract void setProperty(String propertyName,
        Object propertyValue);
6:     public abstract Object getProperty(String propertyName);
7:     public abstract Object[] getProperties();
8:     public abstract String[] getPropertyNames();
9:     public abstract double computeValue(Object complexObject);
10: }
```

Código-fonte 3 – Interface Java para função de similaridade do O-Flm/CO

```
1: package br.oracle.pluginInterfaces;
2: import java.util.Vector;
3: public interface InterfaceSimilarityFunction {
4:     public String getSimilarityName();
5:     public void addExtractor(OFImPlugin oFImPlugin);
6:     public Vector<OFImPlugin> getExtractors();
7:     public double computeSimilarity(double[] vectModel, double
        [] vectSUT);
8:     public double[] getVectorSimilarity(String complexObject);
9: }
```

Código-fonte 4 – Interface Java para extratores em imagens do O-Flm/CO

```
1: package br.oracle.pluginInterfaces;
2: import java.io.IOException;
3: public interface InterfaceImageExtractor {
4:     public abstract String getName();
5:     public abstract void setProperty(String propertyName,
        Object propertyValue);
6:     public abstract Object getProperty(String propertyName);
7:     public abstract Object[] getProperties();
8:     public abstract String[] getPropertyNames();
9:     public abstract double computeValue(Object complexObject)
        throws IOException;
10: }
```

Figura: Interfaces do *framework* OFIm/CO.

Passo 5: Condução de experimentos

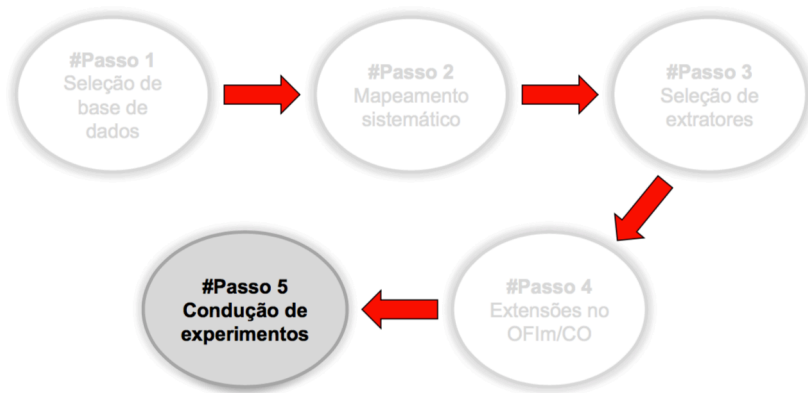


Figura: Fluxo de atividades do mestrado.

Experimentos conduzidos

- Experimento 1: Categorização de imagens sintéticas 3D e 2D de vasos sanguíneos;
- Experimento 2: Avaliação de imagens sintéticas 3D de vasos sanguíneos após aplicação de ruído gaussiano;
- Experimento 3: Oráculos humanos para avaliação da qualidade de sistemas que geram imagens sintéticas 3D de vasos sanguíneos.

Medidas avaliadas

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (1)$$

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (2)$$

$$\text{Especificidade} = \frac{VN}{VN + FP} \quad (3)$$

VP = Verdadeiro Positivo
VN = Verdadeiro Negativo
FP = Falso Positivo
FN = Falso Negativo

Condução

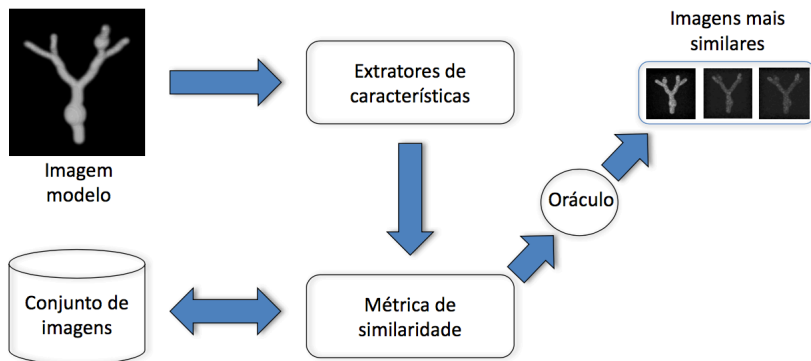


Figura: Fluxo de atividades para julgamento do oráculo.

Experimento 1

- Experimento 1: Categorização de imagens sintéticas 3D e 2D de vasos sanguíneos;
- Experimento 2: Avaliação de imagens sintéticas 3D de vasos sanguíneos após aplicação de ruído gaussiano;
- Experimento 3: Oráculos humanos para avaliação da qualidade de sistemas que geram imagens sintéticas 3D de vasos sanguíneos.

Motivação

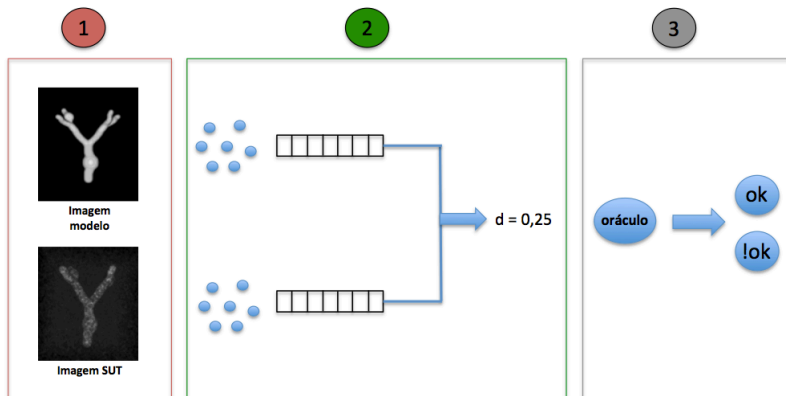


Figura: Abordagem oráculos de teste baseado em características – um passo a passo.

Questões de pesquisa (QPs) e Formulação de hipóteses

- **QP1:** Oráculos de teste baseados em características são capazes de categorizar corretamente imagens sintéticas 3D e 2D de vasos sanguíneos?
- **Hipótese nula H_0 (QP1):** Oráculos de teste baseados em características não são capazes de tomar a decisão correta entre a similaridade das imagens a serem avaliadas a partir de uma imagem modelo.

$$H_0 : \theta_o = 0,05$$

- **Hipótese alternativa H_1 (QP1):** Oráculos de teste baseados em características são capazes de tomar a decisão correta entre a similaridade das imagens a serem avaliadas a partir de uma imagem modelo.

$$H_1 : \theta_o < 0,05$$

Questões de pesquisa (QPs) e Formulação de hipóteses

- **QP2:** Oráculos de teste baseados em características são capazes de categorizar corretamente imagens sintéticas 3D e 2D de vasos sanguíneos em comparação com classificadores randômicos?
- **Hipótese nula H_0 (QP2):** Não há diferença na precisão para categorizar as imagens sintéticas de vasos sanguíneos entre o oráculo de teste baseado em características e os classificadores randômicos.

$$H_0 : \theta_o = \theta_r$$

- **Hipótese alternativa H_1 (QP2):** O oráculo de teste baseado em características obtém um valor de precisão diferente na categorização de imagens sintéticas de vasos sanguíneos em comparação com classificadores randômicos.

$$H_1 : \theta_o \neq \theta_r$$

Objetivos

- Avaliar a precisão da abordagem na avaliação de imagens sintéticas 3D e 2D de vasos sanguíneos;
- Definir um conjunto de extratores de características para imagens sintéticas 3D e 2D de vasos sanguíneos;
- Comparar a precisão da abordagem em relação a classificadores randômicos;
- Relatar as principais limitações e dificuldades na avaliação de imagens sintéticas 3D e 2D de vasos sanguíneos.

Metodologia

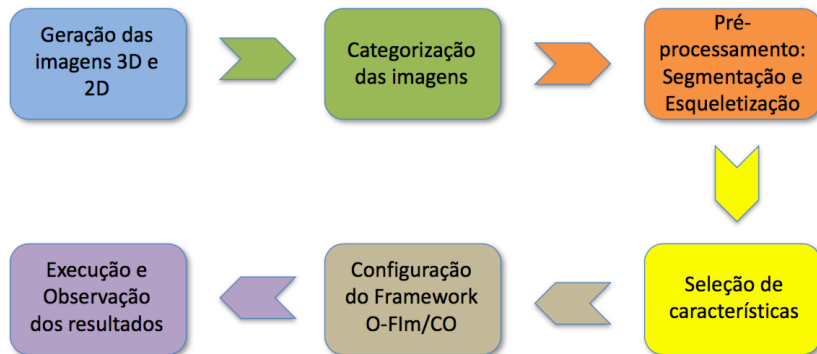


Figura: Etapas para condução do experimento.

Categorização das imagens

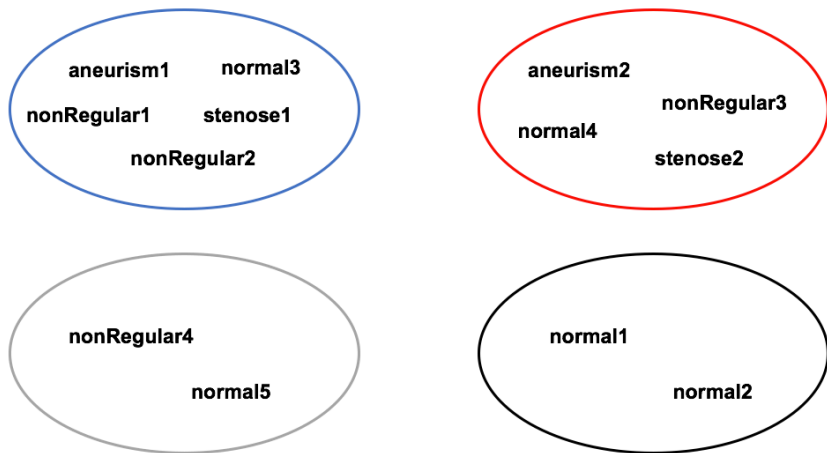
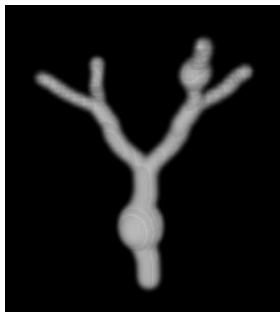


Figura: Categorias de imagens definidas a partir da similaridade entre o número de características – Categorização manual.

Seleção de características



Características extraídas

Dimensões da imagem: 120 120 120

Extrator de densidade

Densidade: $1,03 \times 10^{-4}$

Extrator de bifurcação

Números de bifurcações: 3

Extrator de pontos terminais

Número de pontos terminais: 5

Extrator de segmentos

Número de segmentos: 7

Figura: Exemplo de características obtidas por meio dos extratores desenvolvidos – Estudo 1.

Configuração do limiar

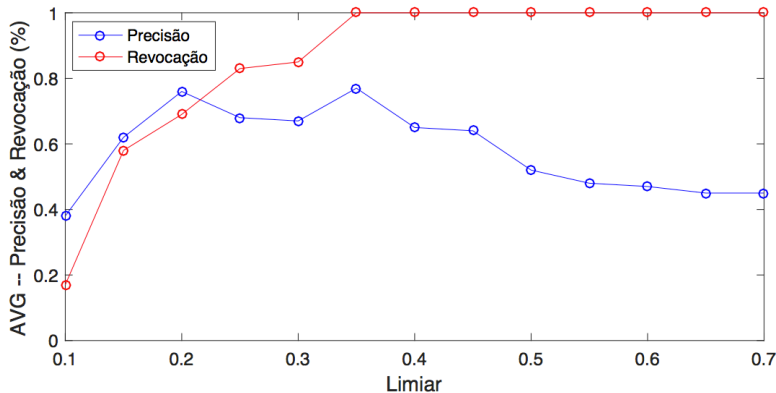


Figura: Comparação da média da precisão e revocação para o conjunto de valores de limiares.

Resultados – QP1

X	VP/VN e FP/FN				Métricas		
Imagem	VP	VN	FP	FN	Precisão	Revocação	Especificidade
aneurism1	4	7	1	0	0,80	1,00	0,87
aneurism2	3	9	0	0	1,00	1,00	1,00
nonRegular1	4	6	2	0	0,67	1,00	0,75
nonRegular2	4	6	2	0	0,67	1,00	0,75
nonRegular3	3	8	1	0	0,75	1,00	0,90
nonRegular4	1	11	0	0	1,00	1,00	1,00
normal1	1	9	2	0	0,33	1,00	0,81
normal2	1	7	4	0	0,20	1,00	0,63
normal3	4	7	1	0	0,80	1,00	0,87
normal4	3	9	0	0	1,00	1,00	1,00
normal5	1	11	0	0	1,00	1,00	1,00
stenose1	4	7	1	0	0,80	1,00	0,87
stenose2	3	9	0	0	1,00	1,00	1,00
Média	—	—	—	—	0,77	1	0,88

Tabela: Resultados retornados pelo oráculo por meio das métricas de precisão, revocação e especificidade para imagens 3D.

Resultados – QP1

X	VP/VN e FP/FN				Métricas		
Imagem	VP	VN	FP	FN	Precisão	Revocação	Especificidade
aneurism1	2	7	1	2	0,67	0,50	0,87
aneurism2	3	9	0	0	1,00	1,00	1,00
nonRegular1	4	7	1	0	0,80	1,00	0,87
nonRegular2	3	7	1	1	0,75	0,75	0,87
nonRegular3	3	9	0	0	1,00	1,00	1,00
nonRegular4	0	11	0	1	0,00	0,00	1,00
normal1	1	11	0	0	1,00	1,00	1,00
normal2	1	7	4	0	0,20	1,00	0,63
normal3	4	7	1	0	0,80	1,00	0,87
normal4	3	9	0	0	1,00	1,00	1,00
normal5	0	11	0	1	0,00	0,00	1,00
stenose1	3	8	0	1	1,00	0,75	1,00
stenose2	3	9	0	0	1,00	1,00	1,00
Média	—	—	—	—	0,72	0,81	0,93

Tabela: Resultados retornados pelo oráculo por meio das métricas de precisão, revocação e especificidade para imagens 2D.

Resultados – QP1

Abordagem	Métrica	O_1	alcances
Imagens 3D	FP	14	A prob. de exatamente 142 de 156 é $p = 1 \times 10^{-12}$
	FN	0	A prob. de exatamente, ou menos do que, 142 de 156 é $p = 0,9$
	VP	36	A prob. de exatamente, ou mais de, 142 de 156 é $p = 1 \times 10^{-12}$
	VN	106	A prob. de exatamente, ou maior que, 142 de 156 é $p = 1 \times 10^{-12}$
Imagens 2D	FP	8	A prob. de exatamente 142 de 156 é $p = 1 \times 10^{-12}$
	FN	6	A prob. de exatamente, ou menos do que, 142 de 156 é $p = 0,9$
	VP	30	A prob. de exatamente, ou mais de, 142 de 156 é $p = 1 \times 10^{-12}$
	VN	112	A prob. de exatamente, ou maior que, 142 de 156 é $p = 1 \times 10^{-12}$
Random 1	FP	24	A prob. de exatamente 108 de 156 é $p = 1 \times 10^{-12}$
	FN	24	A prob. de exatamente, ou menos de, 108 de 156 é $p = 0,9$
	VP	12	A prob. de exatamente, ou mais de, 108 de 156 é $p = 1 \times 10^{-12}$
	VN	96	A prob. de exatamente, ou maior que, 108 de 156 é $p = 1 \times 10^{-12}$
Random 2	FP	28	A prob. de exatamente 100 de 156 é $p = 1 \times 10^{-12}$
	FN	28	A prob. de exatamente, ou menos de, 100 de 156 é $p = 0,9$
	VP	10	A prob. de exatamente, ou mais do que, 100 de 156 é $p = 1 \times 10^{-12}$
	VN	90	A prob. de exatamente, ou maior que, 100 de 156 é $p = 1 \times 10^{-12}$

Tabela: Teste binomial a partir das métricas coletadas no experimento.

Resultados – QP2

t-test bicaudal: oráculos de teste baseados em características versus classificadores randômicos com nível de significância de $\alpha = 0.05$.

p-value = 0,0109

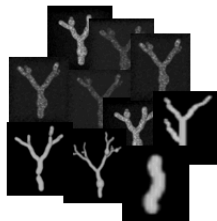
Conclusões

- Oráculos de teste baseados em características são mais eficientes na avaliação da similaridade de imagens sintéticas 3D de vasos sanguíneos;
- A abordagem obteve uma diferença estatisticamente significativa na categorização das imagens 3D e 2D, em comparação a categorização randômica;
- Os resultados obtidos por meio do estudo possibilitaram utilizar subsídios gerados em outros cenários;
- Aplicar a abordagem em outros cenários.

Experimento 2

- Experimento 1: Categorização de imagens sintéticas 3D e 2D de vasos sanguíneos;
- Experimento 2: Avaliação de imagens sintéticas 3D de vasos sanguíneos após aplicação de ruído gaussiano;
- Experimento 3: Oráculos humanos para avaliação da qualidade de sistemas que geram imagens sintéticas 3D de vasos sanguíneos.

Motivação



Questão de pesquisa (QP) e Formulação de hipóteses

- **QP1:** Oráculos de teste baseados em características são capazes de avaliar corretamente a similaridade entre imagens modelo e as respectivas imagens após aplicação de diferentes níveis de ruído gaussiano?
- **Hipótese nula H_0 (QP1):** Oráculos de teste baseados em características não são capazes de tomar a decisão correta entre a similaridade das imagens a serem avaliadas com ruído gaussiano a partir de uma imagem modelo.

$$H_0 : \theta_o = 0,05$$

- **Hipótese alternativa H_1 (QP1):** Oráculos de teste baseados em características são capazes de tomar a decisão correta entre a similaridade das imagens a serem avaliadas com ruído gaussiano a partir de uma imagem modelo com alta precisão.

$$H_1 : \theta_o < 0,05$$

Objetivos

- Avaliar a precisão da abordagem na avaliação de imagens sintéticas 3D após aplicação de ruído gaussiano;
- Definir um conjunto de extratores de características para imagens sintéticas 3D de vasos sanguíneos;
- Avaliar uma abordagem eficaz para definição do limiar para o conjunto de dados;
- Relatar as principais limitações e dificuldades na avaliação de imagens sintéticas 3D de vasos sanguíneos, no contexto do presente experimento.

Metodologia

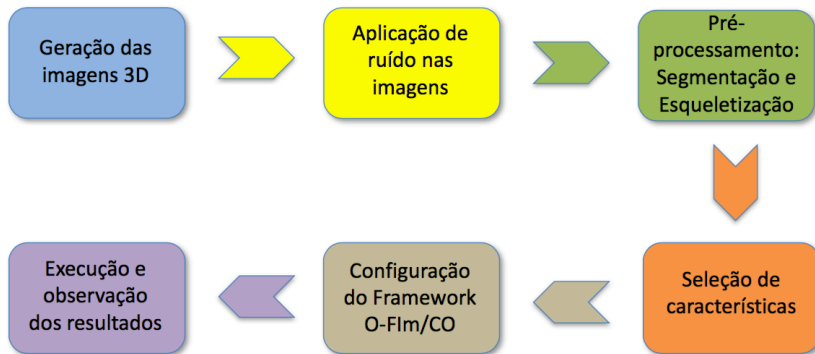
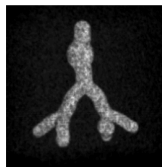


Figura: Etapas para condução do experimento.

Aplicação de ruído gaussiano



sigma = 1



sigma = 2



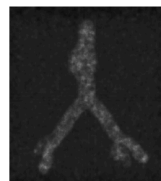
sigma = 3



sigma = 4



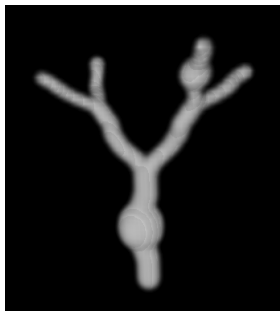
sigma = 5



sigma = 10

Figura: Exemplo de uma imagem de vaso sanguíneo com aneurisma após aplicação de seis diferentes níveis de ruído gaussiano.

Seleção de características



Características extraídas

Dimensões da imagem: 120 120 120

Extrator de densidade

Densidade: $1,03 \times 10^{-4}$

Extrator de bifurcação

Números de bifurcações: 3

Extrator de pontos terminais

Número de pontos terminais: 5

Extrator de segmentos

Número de segmentos: 7

Extrator de tamanho médio de segmentos

Tamanho médio de segmentos: 17,28

Figura: Exemplo de características obtidas por meio dos extratores desenvolvidos.

Configuração do limiar

$$\bar{x} = \frac{\sum_{i=0}^{N-1} x_i}{N} \quad (4)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{x})^2} \quad (5)$$

Assim, os valores definidos para os limiares são apresentados a seguir:

- **th1: 0,313273919**
- **th2: 0,15663696**
- **th3: 0,626547839**

Resultados – QP1

X	VP/VN e FP/FN				Métricas		
Imagem	VP	VN	FP	FN	Precisão	Revocação	Especificidade
aneurism1	6	42	0	0	1,00	1,00	1,00
aneurism2	6	42	0	0	1,00	1,00	1,00
nonRegular1	6	39	3	0	0,66	1,00	0,92
nonRegular2	3	39	3	3	0,5	0,50	0,92
normal1	4	42	0	2	1,00	0,66	1,00
normal2	4	42	0	2	1,00	0,66	1,00
stenose1	6	42	0	0	1,00	1,00	1,00
stenose2	6	42	0	0	1,00	1,00	1,00
Média	—	—	—	—	0,89	0,85	0,98

Tabela: Desempenho do oráculo por meio das métricas de precisão, revocação e especificidade, com o primeiro conjunto de características – Limiar = *th2*.

Resultados – QP1

X	VP/VN e FP/FN				Métricas		
Imagem	VP	VN	FP	FN	Precisão	Revocação	Especificidade
aneurism1	6	42	0	0	1,00	1,00	1,00
aneurism2	6	42	0	0	1,00	1,00	1,00
nonRegular1	6	42	0	0	1,00	1,00	1,00
nonRegular2	3	42	0	3	1,00	0,5	1,00
normal1	4	42	0	2	1,00	0,66	1,00
normal2	4	42	0	2	1,00	0,66	1,00
stenose1	6	42	0	0	1,00	1,00	1,00
stenose2	6	42	0	0	1,00	1,00	1,00
Média	—	—	—	—	1,00	0,85	1,00

Tabela: Desempenho do oráculo por meio das métricas de precisão, revocação e especificidade, com o segundo conjunto de características – Limiar = *th2*.

Resultados – QP1

Limiar	Métrica	O_1	alcances
th1	FP	34	A prob. de exatamente 346 de 384 é $p = 1 \times 10^{-12}$
	FN	4	A prob. de exatamente, ou menos que, 346 de 384 é $p = 0,9$
	VP	44	A prob. de exatamente, ou mais que, 346 de 384 é $p = 1 \times 10^{-12}$
	VN	302	A prob. de exatamente, ou maior que, 346 de 384 é $p = 1 \times 10^{-12}$
th2	FP	0	A prob. de exatamente 377 de 384 é $p = 1 \times 10^{-12}$
	FN	7	A prob. de exatamente, ou menor, 377 de 384 é $p = 0,9$
	VP	41	A prob. de exatamente, ou mais de, 377 de 384 é $p = 1 \times 10^{-12}$
	VN	336	A prob. de exatamente, ou maior que, 377 de 384 é $p = 1 \times 10^{-12}$
th3	FP	133	A prob. de exatamente 251 de 384 é $p = 1 \times 10^{-12}$
	FN	0	A prob. de exatamente, ou menor que, 251 de 384 é $p = 0,9$
	VP	48	A prob. de exatamente, ou mais que, 251 de 384 é $p = 1 \times 10^{-12}$
	VN	203	A prob. de exatamente, ou maior que, 251 de 384 é $p = 1 \times 10^{-12}$

Tabela: Teste binomial a partir das métricas coletadas no experimento.

Conclusões

- A abordagem apresentou alta precisão no julgamento correto das imagens sintéticas 3D de vasos sanguíneos após aplicação de diferentes níveis de ruído gaussiano;
- A alta precisão do oráculo está relacionada a alguns fatores observados na condução do estudo como, por exemplo:
 - O número de características utilizadas refletem diretamente na precisão do oráculo;
 - Características que se complementam tornam o oráculo mais preciso;
 - Abordagens de definição de limiar baseadas no desvio padrão produzem melhores resultados.
- Oráculos de teste baseados em características podem ser utilizados como uma estratégia promissora para automatizar a atividade de avaliação de objetos complexos.

Experimento 3

- Experimento 1: Categorização de imagens sintéticas 3D e 2D de vasos sanguíneos;
- Experimento 2: Avaliação de imagens sintéticas 3D de vasos sanguíneos após aplicação de ruído gaussiano;
- Experimento 3: Oráculos humanos para avaliação da qualidade de sistemas que geram imagens sintéticas 3D de vasos sanguíneos.

Motivação

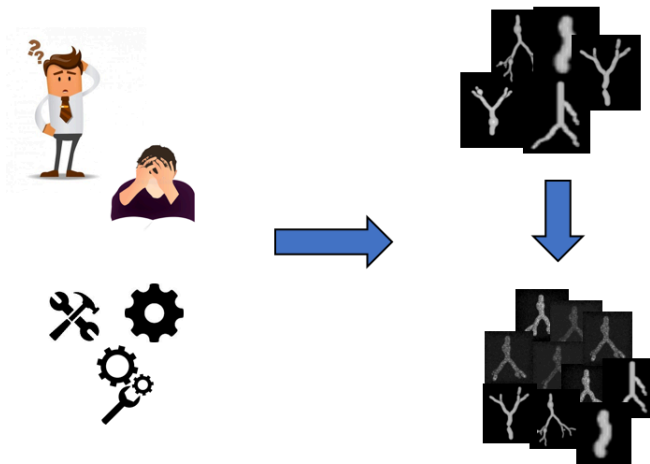


Figura: Avaliação de imagens com ruído gaussiano: oráculos humanos e oráculos automatizados.

Questão de pesquisa (QP) e Formulação de hipóteses

- **QP1:** Oráculos humanos apresentam alta precisão no julgamento correto da similaridade entre imagens modelo e as respectivas imagens após aplicação de ruído em relação a oráculos de teste baseados em características?
- **Hipótese nula H_0 (QP1):** Oráculos de teste baseados em características (θ_a) e oráculos humanos (θ_h) apresentam equivalência no julgamento correto da similaridade das imagens modelo com as respectivas imagens após aplicação de ruído gaussiano.

$$H_0 : \theta_a = \theta_h$$

- **Hipótese alternativa H_1 (QP1):** Oráculos de teste baseados em característica (θ_a) e oráculos humanos (θ_h) não apresentam equivalência no julgamento correto da similaridade das imagens modelo com as respectivas imagens após aplicação de ruído gaussiano.

$$H_1 : \theta_a \neq \theta_h$$

Objetivos

- avaliar a eficiência dos oráculos humanos no julgamento correto da similaridade entre as imagens modelo com as imagens a serem avaliadas em comparação a oráculos automatizados.
- replicar o experimento 2, visando verificar a eficiência dos oráculos de teste baseados em características em comparação aos oráculos humanos;
- realizar avaliações estatísticas objetivando comparar o desempenho do julgamento do oráculo de teste baseado em características em relação aos oráculos humanos;
- relatar as principais limitações e dificuldades na avaliação de imagens sintéticas 3D de vasos sanguíneos, no contexto do presente experimento.

Metodologia

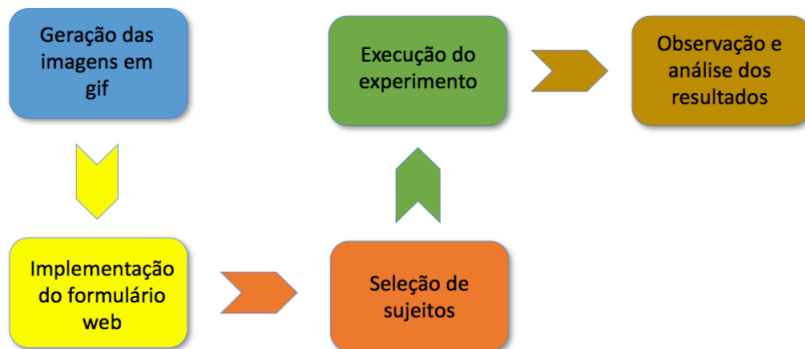


Figura: Etapas para condução do experimento.

Formulário de pesquisa

Oráculo humano para avaliação da qualidade de sistemas que geram imagens médicas sintéticas de redes vasculares

Objetivo: O objetivo desta pesquisa é determinar os critérios e a percepção dos indivíduos humanos quanto à qualidade das imagens de imagens médicas sintéticas de redes vasculares após aplicação de redes de redes generativas.

Toda imagem de rede é avaliada a partir de um sistema que produz imagens médicas sintéticas de redes vasculares, no formato de rede de software. O participante visualiza as imagens e a percepção de cada objeto de avaliação na identificação visual. Informações pessoais não serão fornecidas para qualquer sistema de rede para uso independente.

Orientações

Use a imagem modelo para entender melhor a qualidade e a aparência das imagens que serão avaliadas a seguir. Para facilitar a avaliação das imagens, as imagens foram geradas em formato .png, permitindo uma visualização em 3D.

Observe que as imagens foram geradas em 3D por um tempo determinado. Caso precise olhar alguma imagem com mais detalhes e por mais tempo, clique com o botão direito em cima da imagem e selecione para abrir em uma nova aba de sua navegador.

É aconselhado o uso das navegadores Google Chrome ou Safari, devido a performance observada nos testes realizados para visualização de imagens em formato .png.

Nome *

Your answer

Exemplo

Adapte o formato das imagens anteriores, uma gerada como modelo, sem ruído, e outra imagem com ruído. Note que a presença do ruído na imagem prejudica sua aparência, embora as imagens apresentem características similares. Por isso, nesse exemplo, julgue a similaridade das imagens que serão apresentadas nas próximas páginas deste formulário.

Imagem modelo

Imagem Modelo 1

Marque as imagens que, ao seu ver, são mais similares a imagem apresentada:

Origem??

Origem para sua participação durante o preenchimento deste formulário de pesquisa. Prossiga de mais alguma dificuldade para registro de sua participação (?)

Horário de término (exemplo: 8:30) *

AM PM

Quais dificuldades você teve para analisar a similaridade das imagens? *

- ☐ A quantidade das imagens está muito ruim (tamanho, iluminação, cor, etc).
- ☐ É muito cansativo avaliar as imagens manualmente.
- ☐ É difícil encontrar um padrão nas imagens que facilite identificar a similaridade.
- ☐ Outra.

BACK **SUBMIT**

Figura: Screenshots das etapas de preenchimento do formulário.

Resultados – QP1

X	Oráculos humanos			Oráculo automatizado		
	Prec.	Rev.	Espec.	Prec.	Rev.	Espec.
aneurism1	0,68961	0,72222	0,93015	1,00	1,00	1,00
aneurism2	0,60749	0,67777	0,92010	1,00	1,00	1,00
nonRegular1	0,47701	0,70740	0,88941	1,00	1,00	1,00
nonRegular2	0,78791	0,73703	0,96031	1,00	0,50	1,00
normal1	0,97804	0,87936	0,99629	1,00	0,66	1,00
normal2	0,87388	0,72962	0,98465	1,00	0,66	1,00
stenose1	0,76248	0,65925	0,95608	1,00	1,00	1,00
stenose2	0,67966	0,67777	0,93492	1,00	1,00	1,00
Média	0,73201	0,72380	0,94648	1,00	0,85	1,00

Tabela: Desempenho dos oráculos humanos em relação ao oráculo automatizado (*th2*).

Resultados – QP1

Métrica	<i>p-value</i>	<i>t-value</i>	H_0
Precisão	0.0018	4.8671	Rej.
Revocação	0.2000	1.4151	Ac.
Especificidade	0.0035	4.3245	Rej.

Tabela: *T*-test entre o oráculo de teste baseado em características e os oráculos humanos.

Conclusões

- O oráculo de teste baseado em características apresentou equivalência aos oráculos humanos em um cenário (revocação);
- O estudo evidenciou uma diferença estatística em dois cenários (precisão e especificidade);
- Oráculos de teste baseados em características podem complementar oráculos humanos em atividades de avaliação de objetos complexos;
- Três particularidades devem ser ressaltadas: (i) a quantidade de objetos a serem analisados interfere no julgamento dos oráculos humanos; (ii) o perfil dos participantes (oráculos humanos) interfere nos resultados do experimento.

Conclusões

Contribuições

- Avaliações experimentais da abordagem em sistemas que gerem ou processem imagens médicas 3D, destacando suas vantagens e desvantagens;
- Extensões e configurações realizadas na arquitetura do O-Flm/CO como suporte adequado aos oráculos de teste baseados em características;
- Catálogo de *plug-ins* que representam extratores de características para imagens médicas 3D de vasos sanguíneos foram disponibilizados de modo *open-source*;
- Um MS sobre abordagens e estratégias de teste de software e análise de qualidade em sistemas biomédicos.

Trabalhos futuros

- Aplicação da abordagem em sistemas que gerem ou processem imagens médicas reais – ICMC/USP e IME/USP (**em planejamento**);
- Avaliação de técnicas de geração de escoamento de fluido – ICMC/USP e EACH/USP (**em planejamento**);
- Avaliação da aplicabilidade de métodos de aprendizagem de máquina na abordagem – ICMC/USP, UFV e UFPR (**em andamento**);
- Avaliação da aplicabilidade de teste metamórfico – ICMC/USP (**em aberto**);
- Implementação de *plugins* para o O-Flm/CO – ICMC/USP (**em aberto**).

Publicações

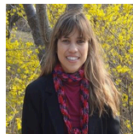
Publicações

- JUNIOR, MISAEL C.; DELAMARO, M. E. Automatização de oráculos de teste para o processamento de imagens médicas de modelos tridimensionais. In: Congresso Brasileiro de Software: Teoria e Prática (CBSOFT), 2016, Maringá, PR. VI WORKSHOP DE TESES E DISSERTAÇÕES DO CBSOFT (WTDSOFT 2016), 2016. p. 35-41. (publicado);
- JÚNIOR, M. C.; OLIVEIRA, R. A. P. ; VALVERDE, M. A. G. ; JACKOWSKI, M. P. ; NUNES, F. L. S. ; DELAMARO, M. E. . Feature-Based Test Oracles to Categorize Synthetic 3D and 2D Images of Blood Vessels. In: the 2nd Brazilian Symposium, 2017, Fortaleza. Proceedings of the 2nd Brazilian Symposium on Systematic and Automated Software Testing - SAST. New York: ACM Press, 2017. p. 1. (publicado);
- JÚNIOR, M. C.; OLIVEIRA, R. A. P. ; VALVERDE, M. A. G. ; JACKOWSKI, M. P. ; NUNES, F. L. S. ; DELAMARO, M. E. . Software testing and quality analysis in biomedical systems: a systematic mapping study. (em revisão);
- JÚNIOR, M. C.; OLIVEIRA, R. A. P. ; VALVERDE, M. A. G. ; JACKOWSKI, M. P. ; NUNES, F. L. S. ; DELAMARO, M. E. . Feature-Based Test Oracles to evaluate three-dimensional synthetic images of blood vessels: a study case. (em andamento)

Agradecimentos



Prof. Dr. Márcio Delamaro



Prof.ª Dr.ª Fátima Nunes



Prof. Dr. Rafael Oliveira



Dr. Miguel Galarreta



Prof. Dr. Marcel Jackowski



Referências I

- BARR, E. T., HARMAN, M., MCMINN, P., SHAHBAZ, M., and YOO, S. (2015). The oracle problem in software testing: A survey. *IEEE Transactions on Software Engineering*, pages 507–525.
- DELAMARO, M. E., NUNES, F. L. S., and OLIVEIRA, R. A. P. (2013). Using concepts of content-based image retrieval to implement graphical testing oracles. *Software Testing, Verification and Reliability*, pages 171–198.
- FILHO, A. C. S., RODRIGUES, E. P., Junior, J. E., and CARNEIRO, A. A. O. (2014). A computational tool as support in b-mode ultrasound diagnostic quality control. *Revista Brasileira de Engenharia Biomédica*, pages 402–405.
- GALARRETA-VALVERDE, M. A., MACEDO, M. M. G., MEKKAAOI, C., and JACKOWSKI, M. P. (2013). Three-dimensional synthetic blood vessel generation using stochastic L-systems. In *Proceedings of the Medical Imaging: Image Processing*, pages 866911–866911–6.
- GIBSON, N. M., DUDLEY, N. J., and GRIFFITH, K. (2001). A computerised quality control testing system for b-mode ultrasound. *Ultrasound in medicine & biology*, pages 1697–1711.
- KITCHENHAM, B. (2004). Procedures for performing systematic reviews. *Keele, UK*, pages 1–26.
- OLIVEIRA, R. A. P. (2012). Apoio à automatização de oráculos de teste para programas com interfaces gráficas. Master's thesis, Instituto de Ciências Matemáticas e de Computação (ICMC) – Universidade de São Paulo (USP), São Carlos, SP.
- OLIVEIRA, R. A. P. (2017). *Test oracles for systems with complex outputs: the case of TTS systems*. Ph.d. thesis, Instituto de Ciências Matemáticas e de Computação (ICMC) – Universidade de São Paulo (USP), São Carlos, SP.
- OLIVEIRA, R. A. P., Gil, V. N., Nunes, F. L. S., and Delamaro, M. E. (2014). An extensible framework to implement test oracle for “non-testable programs”. In *Proceedings of the 26th International Conference on Software Engineering & Knowledge Engineering*, pages 199–204.
- OLIVEIRA, R. A. P., KANEWALA, U., and NARDI, P. A. (2014a). Automated test oracles: State of the art, taxonomies, and trends. *Advances in Computers*, v. 95, pages 113–199. Elsevier.
- OLIVEIRA, R. A. P., MEMON, A. M., GIL, V. N., NUNES, F. L. S., and DELAMARO, M. E. (2014b). An extensible framework to implement test oracles for non-testable programs. In *Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering (SEKE)*, pages 199–204.
- WEYUKER, E. J. (1982). On testing non-testable programs. *The Computer Journal*, pages 465–470.

Obrigado!

Misael Costa Júnior - misaeljr@usp.br

