Avaliação de Classificadores de Imagem Multi-rótulo

por Igor Goulart Cabral

Programa de Pós-graduação em informática
Pontifícia Universidade Católica de Minas Gerais

Objetivos da Pesquisa

- Explorar a aplicação do LV-CIT em um ambiente black-box
- Replicar e estender os experimentos originais

LV-CIT

- É uma técnica de teste de software que busca encontrar defeitos causados pela interação entre diferentes parâmetros de uma imagem.
- Normalmente estes testes são feitos usando pares de imagens.

LV-CIT

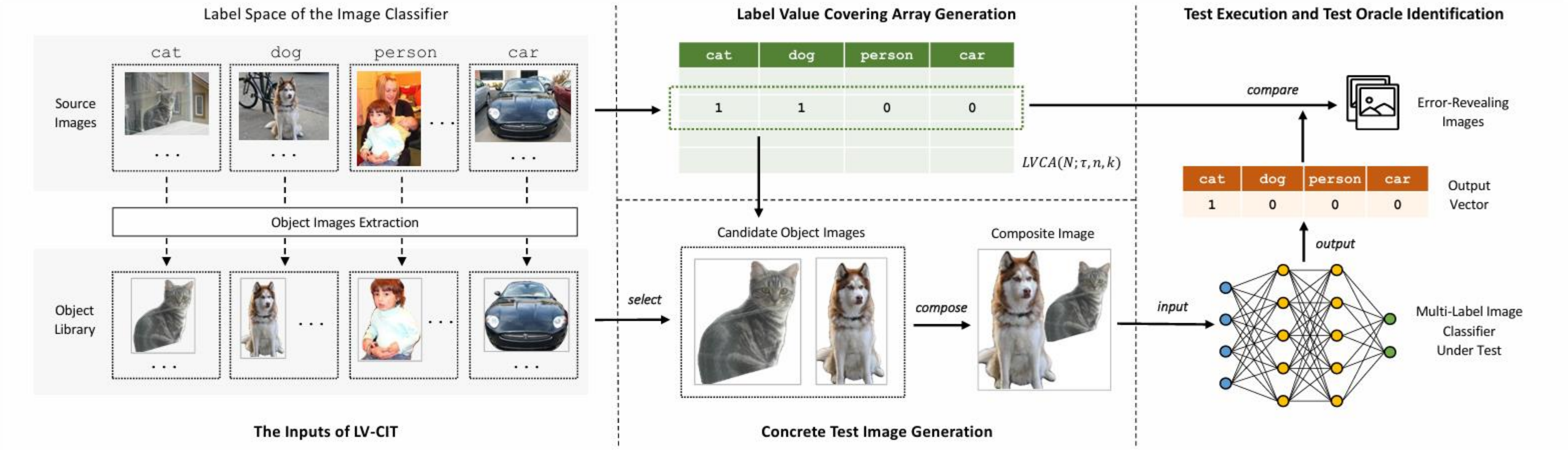


Fig. 1. The workflow of the LV-CIT method.

Metodologia Experimental

Array que mostra a presença do label na imagem

TABLE I
A 2-WAY LABEL VALUE COVERING ARRAY $LVCA(6; 2, 4, 2)$

	l_1 : *cat*	l_2 : *dog*	l_3 : *person*	l_4 : *car*
t_1	1	1	0	0
t_2	1	0	1	0
t_3	1	0	0	1
t_4	0	1	1	0
t_5	0	1	0	1
t_6	0	0	1	1

Metodologia Experimental

Rodar o ambiente LV-CIT

Carregar as bibliotecas e dependências

Carregar os modelos de redes neurais profundas
E os datasets

Testar a acurácia dos modelos com teste blackbox

Algorithm 1 The Adaptive Sampling Algorithm for τ-way Label Value Covering Array Generation

Input: Size of label space n, counting constraint variable k, and covering strength τ

Output: $LVCA(N; \tau, n, k)$

$coverage \leftarrow 0, A \leftarrow \{\}, labels \leftarrow [1, 2, \cdots, n]$
while $coverage < 1$ **do**
  $row \leftarrow [0, 0, \cdots, 0]$ ▷ length is n
  $c \leftarrow RandomInt(1, k)$ ▷ number of value one assigned
  Sort $labels$ by the occurrence of value one in A in ascending order
  $m \leftarrow RandomInt(1, \lceil \frac{n}{2} \rceil)$
  $least_m \leftarrow labels[: m]$, $others \leftarrow labels[m :]$
  $c_1 = min(m, \lceil \frac{c}{2} \rceil)$
  $c_2 = c - c_1$
  $selected \leftarrow RandomSample(least_m, c_1)$
  $selected \leftarrow selected \cup RandomSample(others, c_2)$
  for $i \in selected$ **do**
    $row[i] \leftarrow 1$
  end for
  if $row \notin A$ **then**
    $A \leftarrow A \cup \{row\}$
    Update $coverage$ with A, τ
  end if
end while
for $row \in A$ **do**
  Delete row from A if coverage of $A \setminus \{row\}$ is 1
end for
return A

Análise de Resultados

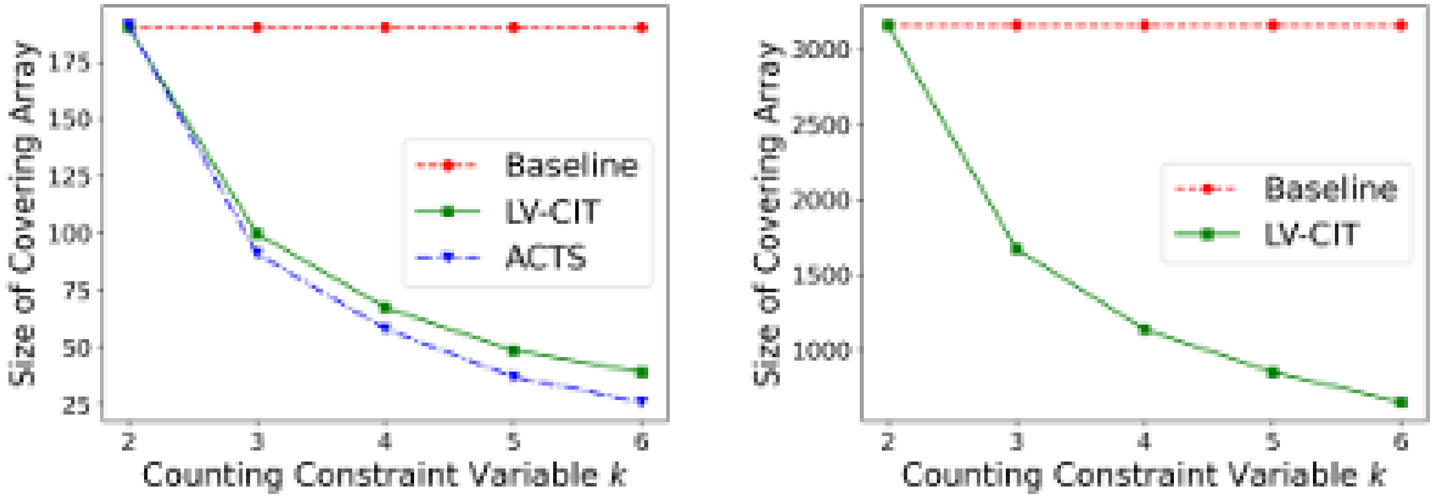
Tabela de validação

TABLE II
THE SIZE OF LABEL SPACES (n), NAME OF DNN MODELS, MEAN AVERAGE PRECISIONS (mAP) ACHIEVED, NUMBER OF IMAGES IN THE OBJECT LIBRARIES (OL), AND NUMBER OF LABELS INVOLVED (LI) OF THE SUBJECT DNN MODELS IN THE EXPERIMENT

Datasets	n	**DNN Models**	mAP	OL	LI
VOC [32]	20	MSRN [45]	96%	580	20
		ML-GCN [22]	94%	576	20
		ASL [46]	94.6%	576	20
COCO [33]	80	MSRN [45]	83.4%	1,973	80
		ML-GCN [22]	83%	1,987	80
		ASL [46]	86.6%	1,811	79

Análise de Resultados

Gráficos casos de teste

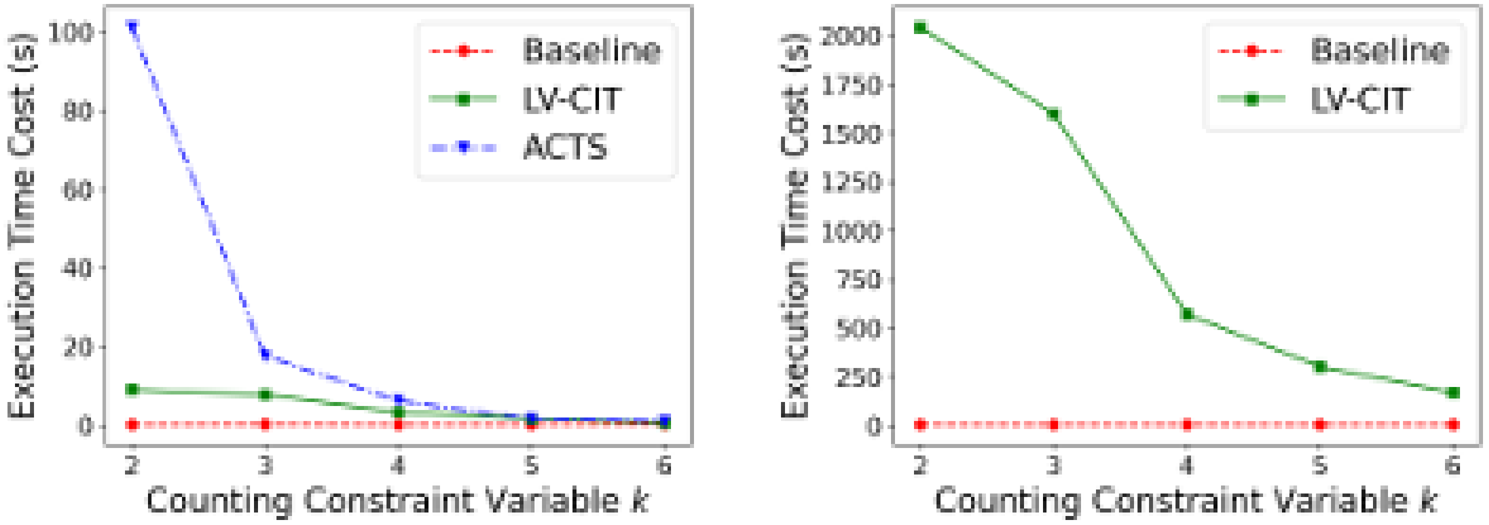


(a) $n = 20$ (b) $n = 80$

Fig. 3. Sizes of 2-way label value covering arrays generated by different generation methods.

Análise de Resultados

Gráficos custo do tempo de execução da cobertura de teste



(a) $n = 20$ (b) $n = 80$

Fig. 4. Execution time costs (seconds) of different generation methods.

Análise de Resultados

Teste de cobertura com LV-CIT e cobertura com método randomico

TABLE III
THE AVERAGE 2-WAY LABEL VALUE COMBINATION COVERAGE OF THE LV-CIT AND *Random* METHODS

Datasets	DNN Models	Coverage	
		LV-CIT	Random
VOC	MSRN	100%	82.34%
	ML-GCN	100%	
	ASL	100%	
COCO	MSRN	100%	90.07%
	ML-GCN	100%	
	ASL	98.75%	

Análise de Resultados

Número de imagens, Erros e entre LV-CIT, Método randomico e ATOM.

TABLE IV
THE NUMBERS OF TEST IMAGES GENERATED, NUMBER OF ERRORS REVEALED, AND PROPORTION OF ERROR-REVEALING IMAGES OF THE LV-CIT, *Random*, AND *ATOM* METHODS

Datasets	DNN Models	Number of Images			Number of Errors			Proportion of Error-Revealing Images		
		LV-CIT	Random	ATOM	LV-CIT	Random	ATOM	LV-CIT	Random	ATOM
VOC	MSRN	674	674	898	603.6	160.8	304	89.6%	23.85%	33.85%
	ML-GCN	674			624.2	169.8	302	92.64%	25.22%	33.63%
	ASL	674			605.4	132	193	89.84%	19.58%	21.49%
COCO	MSRN	11,358	11,358	13,297	10,368.8	6,255.8	5,984	91.3%	55.08%	45%
	ML-GCN	11,358			10,387	6,291.2	6,484	91.46%	55.39%	48.76%
	ASL	11,352			9,915	5,645.2	4,584	87.35%	49.71%	34.47%

Análise de Resultados

Precisão de reconhecimento e interação de rótulos

TABLE V

THE MEAN AVERAGE PRECISION (MAP) AND 2-WAY MEAN INTERACTION ACCURACY (MIA) OF THE SUBJECT DNN MODELS

Datasets	DNN Models	*mAP*	2-way *mIA*
VOC	MSRN	96%	81.79%
	ML-GCN	94%	80.28%
	ASL	94.6%	84%
COCO	MSRN	83.4%	95.51%
	ML-GCN	83%	95.53%
	ASL	86.6%	96.01%

Análise de Resultados

3 principais combinações de rótulos que podem gerar Erros

TABLE VI
THE NUMBER OF ERRORS-REVEALING IMAGES (E), AND NUMBER OF RELATED TRAINING IMAGES (TI) OF TOP THREE 2-WAY LABEL COMBINATIONS (LC) THAT CAUSE THE MOST ERRORS ACROSS THE THREE TYPES OF ERRORS BY THE ALS IMAGE CLASSIFIER

Datasets	Missing			Extra			Mismatch		
	LC	E	TI	LC	E	TI	LC	E	TI
VOC	{chair, cow}	28.8	0	{car, person}	3.6	215	{bus, car}	8	63
	{bus, chair}	22	0	{horse, person}	1.8	221	{car, chair}	2.2	5
	{dining table, train}	21.8	0	{motorbike, person}	1.8	169	{dining table, person}	2	94
COCO	{donut, surfboard}	32	1	{bus, person}	26	2,149	{knife, skis}	34	1
	{kite, skateboard}	30	2	{baseball glove, sports ball}	23.4	826	{cake, dining table}	27	1,234
	{couch, spoon}	28.8	74	{keyboard, mouse}	22.8	878	{cow, dog}	26.6	49

Conclusões e Trabalhos Futuros

- [ ] Implementar o modelo
- [ ] Ampliar os usos para novos datasets e novos modelos