

DETECÇÃO DE PEDESTRES USANDO *DEEP LEARNING* EM VANTS

Vieira, Fernando A. Zarate¹ (f02_@hotmail.com.br); Amorim, Willian Paraguassu² (willianamorim@ufgd.edu.br);

¹Discente do curso de Engenharia de Computação da Universidade Federal da Grande Dourados – UFGD;

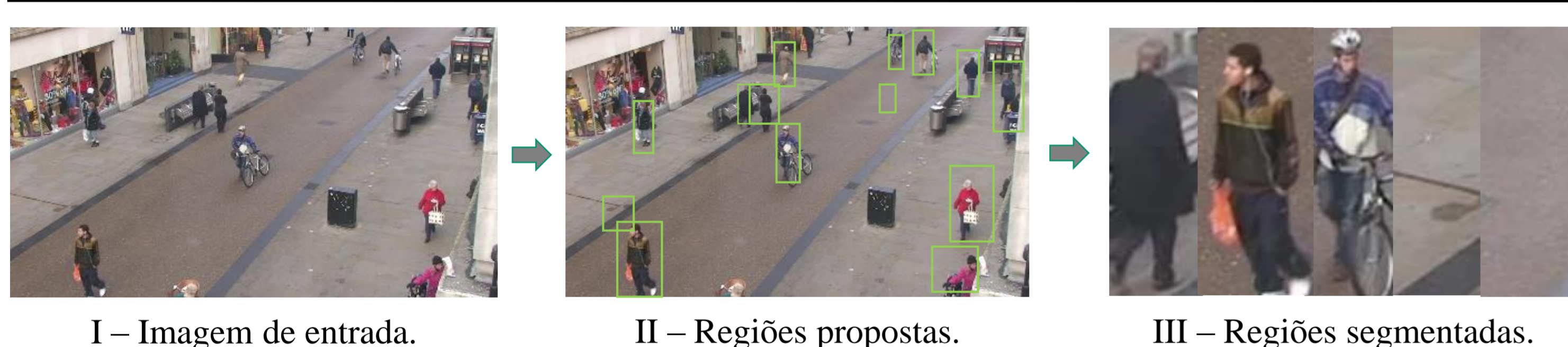
²Docente do curso de Engenharia de Computação da Universidade Federal da Grande Dourados – UFGD

Introdução

A detecção de pedestres é uma tarefa complexa e importante para diversas aplicações, tais como, direção autônoma de veículos, robótica, entretenimento, vigilância, entre outros. Recentemente foi constatado a eficácia de técnicas que envolvem a Aprendizagem Profunda (*Deep Learning*) em diversas tarefas de Inteligência Artificial.

O objetivo desse trabalho é a detecção de pedestres em imagens de vídeo para monitoramento. Esta tarefa foi dividida em duas etapas conforme ilustrado na Figura 1.

DETECÇÃO



CLASSIFICAÇÃO

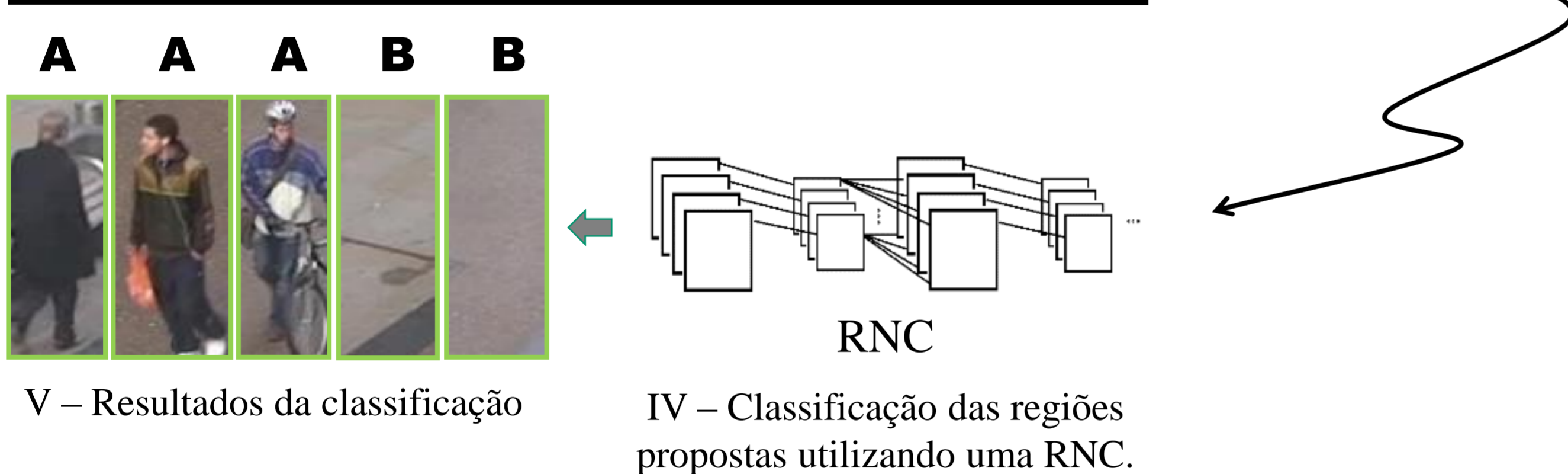


Figura 1. Ilustração do processo de detecção e classificação de pedestres.

Metodologia

Neste trabalho, foi utilizado a base de dados *Town Centre*, devido a quantidade de trabalhos publicados para o problema de detecção de pedestres e possuir um posicionamento de filmagem semelhante ao de um Veículo Aéreo Não Tripulado (VANT). A base de dados consiste em um vídeo de vigilância com duração de cinco minutos. Através de uma rotina, foram extraídas 7.928 amostras, sendo 1.992 de não pedestre e 5.936 de pedestre, além de outras informações pertinentes a cada quadro de vídeo. Portanto, foram avaliadas seis estratégias diferentes de treinamento, sendo *fine tuning* com 25%, 50%, 75% e 100%, e configurações com e sem *transfer learning*.

Resultados

A Tabela I apresenta os valores dos resultados para cada estratégia proposta, onde TL representa *Transfer Learning* e FT *Fine Tuning*.

Tabela 1. Métricas de avaliação do treinamento das arquiteturas de Aprendizagem Profunda.

Arquitetura	Estratégia	Tempo (s)	Acurácia	Erro
Inception-V3	Sem TL	7.957,94	98,61%	0.0448
	Com TL	6.298,55	96,73%	0.0890
	25% FT	6.397,55	99,11%	0.0292
	50% FT	6.440,56	99,58%	0.0141
	75% FT	6.480,03	99,73%	0.0119
	100% FT	7.941,78	99,81%	0.0118
ResNet-50	Sem TL	9.229,72	98,73%	0.0464
	Com TL	6.295,54	98,44%	0.0504
	25% FT	6.348,18	99,60%	0.0173
	50% FT	6.378,60	99,75%	0.0128
	75% FT	7.027,34	99,83%	0.0080
	100% FT	9.187,64	99,79%	0.0085
VGG-16	Sem TL	9.864,47	93,09%	0.1713
	Com TL	6.208,76	95,81%	0.1114
	25% FT	6.205,13	99,26%	0.0264
	50% FT	6.496,68	99,56%	0.0196
	75% FT	8.339,03	99,73%	0.0128
	100% FT	9.761,92	99,68%	0.0147
VGG-19	Sem TL	11.321,66	92,94%	0.1752
	Com TL	6.168,48	94,78%	0.1300
	25% FT	6.207,14	99,01%	0.0324
	50% FT	7.103,37	99,62%	0.0177
	75% FT	9.503,26	99,64%	0.0154
	100% FT	11.385,83	99,70%	0.0142
Xception	Sem TL	13.758,18	99,34%	0.0225
	Com TL	6.325,11	97,67%	0.0610
	25% FT	6.417,04	99,28%	0.0251
	50% FT	7.525,73	99,41%	0.0197
	75% FT	8.750,17	99,58%	0.0172
	100% FT	13.796,28	99,51%	0.0207

Conclusão

Neste trabalho foi apresentada uma abordagem proposta para a tarefa de detecção e classificação de pedestres, utilizando a base de imagens *Town Centre*, 5 arquiteturas e duas abordagens de treinamento *FT* e *TL*. Todas as estratégias obtiveram uma taxa de acurácia acima de 92.94%, com destaque para a arquitetura ResNet-50 usando *FT* com 75% dos pesos aplicados a arquitetura, alcançando uma acurácia de 99.83%. Trabalhos futuros visam o estudo de técnicas que resolvam os problemas de detecção automática, problema de oclusão e avaliação de outras técnicas de aprendizagem profundo.

Realização:

UFGD
Universidade Federal
da Grande Dourados

UEMS
Universidade Estadual
de Mato Grosso do Sul

Parceiros:

CAPES

CNPq
Conselho Nacional de Desenvolvimento
Científico e Tecnológico

