ESTUDO COMPARATIVO DE ARQUITETURAS DE REDES NEURAIS EM ROBÔS HUMANOIDES

Jonas Henrique Renolfi de Oliveira[1], Reinaldo Augusto da Costa Bianchi[1]
[1] Centro Universitário FEI, São Bernardo do Campo, SP, Brasil
uniejonoliveira@fei.edu.br, rbianchi@fei.edu.br

Resumo: Neste trabalho, investigamos o desempenho de um sistema de visão para detecção de bola, com base em diferentes configurações da arquitetura da Rede Neural Convolucional MobileNet, em um cenário de hardware restrito. Ao reduzir gradualmente o tamanho da entrada e o número de parâmetros que compõem a rede neural e comparando seu tempo de inferência em um mini PC Intel NUC Core i7, incorporado a um robô humanoide, descobrimos valores aceitáveis para os multiplicadores de largura e resolução a serem usados em nosso sistema de detecção de bolas de futebol, durante uma partida de futebol de robôs.

1. Introdução

É comum que as redes neurais profundas (DNN) sejam programadas para serem executadas de maneira ideal em computadores com unidades de processamento gráfico (GPU). No entanto, em cenários em que o hardware é limitado, como em robôs móveis, o processamento é feito principalmente apenas em CPUs, o que dificulta o desempenho de sistemas baseados em técnicas de aprendizado profundo.

Neste trabalho, é proposto um sistema de visão computacional para robôs móveis baseado em Redes Neurais Convolucionais (CNN), voltado para arquiteturas somente de CPU. O principal objetivo do projeto é criar um algoritmo de detecção de bola que funcione com uma velocidade compatível com a dinâmica de um jogo de futebol de robôs humanoides, onde os robôs estão equipados com um mini-PC Intel NUC Core i7.

O sistema de visão proposto emprega a arquitetura MobileNet [1], uma CNN desenvolvida para sistemas móveis. Para alcançar desempenho próximo ao tempo real, foi realizado um estudo comparativo entre diferentes versões da MobileNet, cuja complexidade é reduzida por meio de seus multiplicadores de largura e resolução. Na próxima seção apresentaremos a metodologia adotada para realizar este estudo comparativo.

A partir de 2012, muitas equipes que participam de competições de classificação e reconhecimento de padrões, como a ILSVRC (Imagenet Large Scale Visual Recognition Challenge) [2], passaram a utilizar redes neurais profundas, melhorando significativamente seus desempenhos.

MobileNets são modelos de Rede Neural Convolucional criados para executar tarefas de visão computacional em sistemas móveis ou embarcados. Eles empregam convoluções separáveis em profundidade [3], uma versão fatorada da operação de convolução que permite a personalização do tamanho do modelo por meio de dois hiperparâmetros: os multiplicadores de largura e de resolução.

O multiplicador de largura altera o número de canais de entrada em cada camada, o número de parâmetros do modelo, e seu custo computacional. O multiplicador de resolução define a resolução da imagem de entrada da rede, reduzindo apenas o custo computacional [1].

2. Metodologia

Esta seção apresenta as configurações de rede e os procedimentos de treinamento, os robôs humanoides usados nos experimentos e o conjunto de dados de imagens da bola de futebol, criado pelo autor da pesquisa, e usado durante a fase de treinamento.

2.1. Arquiteturas de rede e treinamento

Oito configurações da MobileNet foram testadas modificando os valores dos multiplicadores de largura e resolução. Para o multiplicador de largura, foram utilizados os valores 1, 0.75, 0.5 e 0.25, o que resultará em redes com parâmetros aprendíveis de 4, 2, 2.6, 1.3 e 0,5 milhões, respectivamente. Os valores usados para o multiplicador de resolução foram 224 e 128. Os valores combinados de ambos os hiperparâmetros resultaram em um total de oito modelos que foram treinados usando um conjunto de 4400 imagens.

Cada um dos oito modelos foi treinado por um total de oito horas em um servidor com um processador Intel Xeon Gold 5118@2,3 GHz com 48 CPUs, 192 GB de RAM e duas NVIDIA Tesla V100-PCIE com 16 GB de memória cada, executando o CentOS 7.6.1810.

2.2. Robôs Humanoides

Os experimentos foram realizados nos robôs humanoides da equipe RoboFEI. A equipe possui dois robôs da categoria TeenSize, cada um com 19 Servo motores Dynamixel (MX-64, MX-106 e XM430) e 19 graus de liberdade. Os robôs humanoides empregam câmeras Genius WideCam F100 (Full HD) para captura de imagens e um sensor de orientação CH Robotics UM7.

Tabela I – Tempo de inferência das oito configurações de rede em um Intel NUC.

Resolution multiplier	Width multiplier	FPS	Time (s)	
			μ	σ
224	1.00	11	0.0861	0.0106
	0.75	14	0.0669	0.0086
	0.50	18	0.0536	0.0062
	0.25	22	0.0450	0.0059
128	1.00	11	0.0869	0.0109
	0.75	14	0.0692	0.0082
	0.50	18	0.0540	0.0065
	0.25	22	0.0451	0.0055

Tabela II – Medidas de desempenho das oito configurações de rede em um Intel NUC i7-5557U.

Resolution multiplier	Width multiplier	TP	FP	FN	TN	Precision	Recall	Specificity	F-score	Accuracy	mAP
224	1.00	236	1	14	99	0.9958	0.9440	0.9900	0.9692	0.9571	98.78%
	0.75	219	1	31	100	0.9955	0.8760	0.9901	0.9319	0.9088	93.08%
	0.50	210	1	40	100	0.9953	0.8400	0.9901	0.9111	0.8832	95.38%
	0.25	202	11	48	92	0.9484	0.8080	0.8932	0.8726	0.8329	89.52%
128	1.00	236	0	14	100	1.0000	0.9440	1.0000	0.9712	0.9600	97.29%
	0.75	230	1	20	99	0.9957	0.9200	0.9900	0.9563	0.9400	99.20%
	0.50	214	0	36	100	1.0000	0.8560	1.0000	0.9224	0.8971	96.71%
	0.25	154	27	96	80	0.8508	0.6160	0.7477	0.7146	0.6555	71.52%

3. Resultados

A Tabela I apresenta o tempo de inferência para cada configuração da MobileNet, medida em FPS, fremes por segundo. As experiências foram conduzidas em um Intel NUC i7-5557U. Como pode ser visto, quanto menor for o multiplicador de largura, mais rápido será o tempo de inferência, independentemente do multiplicador de resolução. Embora um resultado de comparação mais completo possa ser visto na Tabela II, que apresenta valores de desempenho para as oito configurações de rede no conjunto de dados de teste.

Os resultados apresentados na Tabela II indicam que podemos considerar o tempo de inferência e a precisão de cada multiplicador de largura. Por um lado, o multiplicador de largura menor, resulta em um tempo de inferência mais rápido com a pior precisão e, por outro lado, o multiplicador de largura maior resulta na melhor precisão com o tempo de inferência mais lento.

A Fig. 1 apresenta um subconjunto das medidas de desempenho, juntamente com o tempo médio de inferência de cada configuração de rede. Considerando a dinâmica do jogo, para esse conjunto de dados de bolas de futebol e a arquitetura da Rede Neural Convolucional, devemos considerar o uso de um multiplicador de largura de 0.75 e um multiplicador de resolução de 128. Isso garante uma precisão interessante com um tempo de inferência aceitável.

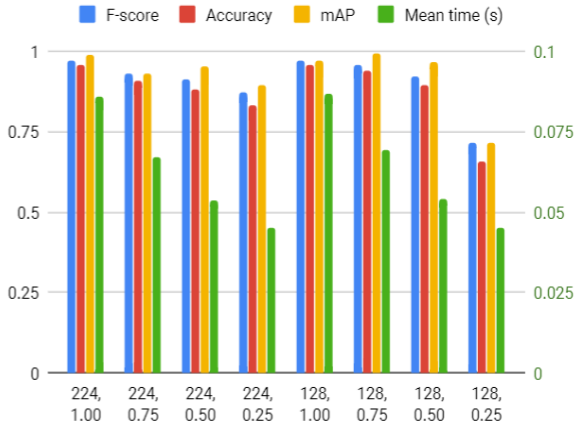

Figura 1 – Medidas de desempenho (eixo esquerdo) e tempo médio de inferência (eixo direito) das oito configurações de rede.

4. Conclusões

Este trabalho apresentou um estudo comparativo do desempenho e tempo de inferência de arquiteturas reduzidas da Rede Neural Convolucional em cenários de hardware restritos para a tarefa de detectar uma bola de futebol no domínio da Liga de Futebol de robôs Humanoides. Oito configurações diferentes do modelo MobileNet foram criadas alterando dois hiperparâmetros do modelo (multiplicador de largura e multiplicador de resolução) e treinados em um conjunto de dados de 4400 imagens anotadas.

Os resultados mostram que todas as configurações da MobileNet atingem tempos de inferência aceitáveis nesse hardware restrito (entre 11 e 22 FPS), tornando o sistema utilizável em tempo real durante partidas de futebol com humanoides. No entanto, as medidas de desempenho variaram bastante, com valores de recall variando de 61.6 % a 94.4 % e uma precisão média (mAP) variando de 71.52 % a 99.2 %. Dadas as descobertas, consideramos o uso de uma configuração com um multiplicador de largura de 0.75 e um multiplicador de resolução de 128, pois resulta em uma arquitetura de rede neural com a menor quantidade de parâmetros aprendíveis, tempo de inferência razoável e valores de desempenho comparáveis aos das melhores configurações que foram testadas.

5. Agradecimentos

À instituição Centro Universitário FEI pela realização das medidas ou empréstimo de equipamentos. Número do projeto FEI: CT-1D1R04/19

6. Referências

[1] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andretto, and H. Adam, “Mobilenets: Efficient convolutionl neural networks for mobile vision aplications,” asXiv prepint arXiv:1704.0486, 2017.

[2]” Imagenet Large scale visual recognition challenge,”2017.

[3] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proccedins of the IEEE conference on computer vision and pattern recognition, pp. 1251-1258, 2017.