



Projeto de Formatura – 2023 – Press Release

PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia Elétrica – Ênfase Computação

Tema:

Resolution-wise Convolutional Neural Networks for Image Classification

Aluno da Escola Politécnica da Universidade de São Paulo desenvolve método para diminuir consumo de recursos computacionais de Redes Neurais Convolucionais

São Paulo, Dezembro, 2023

Em seu projeto de formatura, o aluno Jônatas de Souza Nascimento, orientado pelo professor Artur Jordão, estudou uma abordagem que reduz significativamente o custo computacional das tarefas de classificação de imagens em Redes Neurais Profundas (DNNs), particularmente em Redes Neurais Convolucionais (CNNs), sem sacrificar substancialmente a acurácia. Este desenvolvimento aborda um desafio crucial no campo da visão computacional, onde a crescente profundidade e complexidade das redes neurais têm levado a um aumento nos custos de processamento.

O estudante focou em um dos parâmetros mais influentes que afetam os requisitos computacionais de uma CNN: a resolução espacial da imagem de entrada. Ao empregar um método que transpõe os pesos de redes neurais pré-treinadas para modelos reduzidos com arquiteturas idênticas, o estudo demonstrou com sucesso uma diminuição acentuada nas Operações de Ponto Flutuante (FLOPs) necessárias - uma medida de intensidade computacional, com pouca redução na acurácia.

Seus experimentos foram conduzidos no conjunto de dados ImageNet, utilizando a arquitetura ResNet50, um modelo amplamente utilizado em tarefas de classificação de imagens. Ao reduzir a resolução do modelo de 224x224 pixels para 168x168 pixels, os pesquisadores alcançaram uma redução notável de 29,26% em FLOPs.

Interessantemente, essa diminuição na demanda computacional foi acompanhada apenas por uma queda mínima na acurácia, registrando apenas 2,70%. Alavancado por esses resultados, o trabalho propôs um método que escolhe aleatoriamente a resolução para as redes neurais serem reescaladas de um conjunto pré-definido de resoluções. Esta abordagem reduziu as operações de ponto flutuante em 17,30%, com uma perda de acurácia abaixo de 3%.

Esses desenvolvimentos contribuem positivamente ao campo da visão computacional. Ao possibilitar operações de redes neurais mais eficientes, esta pesquisa pavimenta o caminho para aplicações mais custo-efetivas de tecnologias de classificação de imagens. Mais informações sobre esta pesquisa e suas implicações no campo da IA e visão computacional podem ser encontradas na monografia completa, disponível no site oficial do Departamento de Engenharia de Computação. Este projeto será apresentado nos dias 19 e 20 de dezembro na Escola Politécnica da Universidade de São Paulo.

Aluno: Jônatas de Souza Nascimento

Orientador: Prof. Dr. Artur Jordão