



Evento	Salão UFRGS 2020: SIC - XXXII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2020
Local	Virtual
Título	Aproximação Seletiva de Camadas de uma Rede Neural Convolutacional
Autor	RUBENS LUIZ RECH JÚNIOR
Orientador	PAOLO RECH

Autor: Rubens Luiz Rech Junior
Orientador: Paolo Rech
Instituição: UFRGS

Aproximação Seletiva de Camadas de uma Rede Neural Convolutacional

As novas arquiteturas computacionais de precisão mista permitem que o programador configure a precisão das operações de acordo com as verdadeiras necessidades da aplicação, isto é, de uma maneira mais ajustada em termos de uso dos recursos. Diminuir a precisão reduz tanto o tempo de execução como o consumo de potência do programa, porém também afeta a exatidão dos seus resultados. Faz-se então necessário avaliar cautelosamente quando o uso de operações com precisão reduzida é aceitável para uma determinada aplicação e encontrar estratégias para melhorar o *trade-off* entre ganho de performance e erro introduzido. Nesse trabalho, propomos a aproximação — executar com precisão reduzida — seletiva das camadas de uma rede neural convolutacional (CNN) para detecção de objetos em tempo real. Primeiramente, mostramos que cada camada de uma CNN, quando aproximada, tem um impacto diferente tanto na performance de execução quanto no erro introduzido na saída da rede neural. Sendo assim, avaliamos a relação entre ganho de performance e erro introduzido ao se aproximar seletivamente as camadas de uma CNN. Priorizamos a redução da precisão daquelas camadas que promovem maior ganho de performance, menor erro na saída ou maximizam a relação custo-benefício (performance-erro). Dessa forma, mostramos que nossa estratégia pode acelerar em até 22% a execução, introduzindo um erro médio abaixo de 6% e sem reduzir a acurácia da CNN.