Deep Learning aplicado à estimativa de profundidade em imagens estéreo multivista

SÉRGIO MURILO MOREIRA DE OLIVEIRA[1,4];
ERIKSON FREITAS DE MORAIS[1,2];
MARCELLA SCOCZYNSKI RIBEIRO MARTINS[1,3]

[1.] Universidade Tecnológica Federal do Paraná, Rua Doutor Washington Subtil Chueire, 330, Ponta Grossa -PR, Brasil;
[2.] Orientador, Departamento Acadêmico de Informática, emorais@utfpr.edu.br;
[3.] Coorientador, Departamento Acadêmico de Informática, marcella@utfpr.edu.br;
[4.] Acadêmico, Programa de Pós-Graduação em Ciência da Computação, sergiomurilo@alunos.utfpr.edu.br;

Introdução

A estimativa de profundidade do ambiente é crucial no desenvolvimento de diversas aplicações robóticas e processos de automação. Dentre estes, a automação veicular tem o potencial de mudar a forma como vivemos, aumentando a mobilidade, qualidade de vida e, principalmente, a segurança das pessoas, uma vez que acidentes de trânsito são a oitava maior causa de mortes no mundo.

Atualmente, o LiDAR (*Light Detection and Ranging*) é um dos principais sensores responsáveis pela estimativa de profundidade. Com margem de erro de 2 centímetros a 200 metros do alvo nos modelos mais avançados, fornece informações precisas quanto à profundidade dos objetos nas cenas observadas. Em contrapartida, o alto custo de produção inviabiliza a utilização em larga escala do LiDAR, pois a necessária redundância de sensores em veículos autônomos, por exemplo, os torna comercialmente inviáveis.

Imagens digitais, monoculares e binoculares (captadas a partir de duas câmeras observando a mesma a cena a partir de pontos de vista ligeiramente distintos), apresentam indicadores visuais (*depth cues*). A aplicação de técnicas de Visão Computacional no processamento de tais imagens permite inferir a profundidade da cena observada. Ainda que possuam alta resolução, câmeras digitais são mais acessíveis e baratas em comparação ao LiDAR. Por outro lado, o custo computacional, diretamente proporcional à resolução das imagens processadas, e a baixa precisão em comparação ao LiDAR dificultam a utilização de tais técnicas em aplicações comerciais.

Observando tal cenário, o objetivo deste trabalho é desenvolver um sistema que aplica *Deep Learning* na obtenção do mapa de profundidade da cena observada a partir de duas ou mais imagens da mesma cena, captadas de ângulos ligeiramente distintos.

Palavras-chave*: Visão Estéreo Multivista; Automação Veicular; Aprendizado Profundo.*

Metodologia

O presente trabalho é baseado em: artigos científicos e livros referentes a *Deep Learning*, *Convolutional Neural Networks, Siamese Neural Networks* e *Generative Adversarial Networks*; documentação referente à *TensorFlow, Keras e Python, frameworks* e linguagem de programação utilizados no desenvolvimento do mesmo.

Deep Learning: Aprendizado Profundo é uma classe de algoritmos de Aprendizado de Máquina (*Machine Learning*) baseados em Redes Neurais (*Neural Networks*), cujo objetivo é obter informações e representações características de alto nível a partir de uma entrada de dados brutos. O nome tem origem na arquitetura do modelo implementada nos algoritmos, composta por uma cascata profunda de camadas, onde cada camada utiliza a saída da camada anterior como entrada, gerando um modelo de aprendizado incremental hierarquicamente organizado.

Convolutional Neural Networks: Rede Neural Convolucional é um modelo Rede Neural que emprega a convolução em pelo menos uma de suas camadas. Convolução é um tipo de operação matemática, no qual uma matriz de entrada é multiplicada linearmente por um *kernel*, uma pequena matriz, também conhecido como máscara ou matriz de convolução. O tamanho, formato e valores do *kernel* variam de acordo com o efeito desejado. A convolução pode ser utilizada na detecção de bordas, remoção de ruídos, borramento, etc.

Siamese Neural Networks: Rede Neural Siamesa é um modelo de Rede Neural composto por duas ou mais sub-redes "idênticas", ou seja, com o mesmo número de camadas e parâmetros.

Generative Adversarial Neural Networks: Rede Adversária Generativa (ou Geradora) é um modelo de Aprendizado de Máquina não supervisionado, utilizado para descobrir padrões em uma entrada, de forma que o modelo seja capaz de gerar uma saída plausível a partir da entrada proveniente de um *Dataset.*

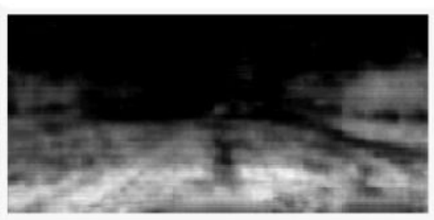
Experimentos e Resultados

O modelo atual consiste em uma Rede Neural Siamesa com 2 sub-redes compostas por camadas convolucionais, responsáveis pelo processamento simultâneo das imagens esquerda e direita. O resultado passa por uma camada de concatenação, sendo utilizado na sequência como semente para uma Rede Adversária Generativa, responsável por gerar o mapa de profundidade da cena observada. Os experimentos atuais foram realizados com base no *DrivingStereo Dataset,* que fornece imagens binoculares provenientes das câmeras esquerda e direita, bem como o mapa de profundidade, obtido a partir de um LiDAR. O resultado dos experimentos obteve 54% de acurácia na geração do mapa de profundidade, podendo ser observado na figura abaixo.

Imagem da câmera esquerda

Imagem da câmera direita

Mapa de profundidade



Resultado obtido pelo modelo

A aplicação de ***Deep Learning*** à **estimativa de profundidade** em **imagens digitais** pode reduzir significativamente os custos de desenvolvimento de diversas **aplicações robóticas**, inclusive de **automação veicular**, contribuindo para melhoria da mobilidade urbana e segurança no trânsito.

UTFPR
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

PPGCC
Programa de Pós-Graduação em Ciência da Computação

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

WPCCG 2021
Workshop de Pesquisa em Computação dos Campos Gerais