Detecção de placas de trânsito em uma imagem

Victor N. Rebli

Universidade Federal do Espirito Santo (UFES) , victor.rebli@gmail.com
Programa de Pós-Graduação em Informática
Av. Fernando Ferrari, No. 514, Prédio CT-VII
Vitória/ES, Brasil, 29075-910

Resumo – Este trabalho apresenta os resultados alcançados em se utilizar redes neurais profundas para dada uma imagem qualquer, detectar a ocorrência de placas de trânsito. Foi realizado alguns experimentos , envolvendo tamanho da base e parâmetros na arquitetura da rede neural e avaliado como tais mudanças afetam a acurácia do modelo.

I. INTRODUÇÃO

O problema de detecção e reconhecimento de objetos em uma imagem é um problema há muito tempo explorado pela comunidade científica e que ao longo dos anos tem gerado boas soluções. Atualmente com as redes neurais profundas, mais precisamente com redes convolucionais, alcançou-se bons nivéis de acurácia em problemas envolvendo classificação de imagens e detecção de objetos. Em especial problemas cujo interesse é detectar objetos em uma imagem geralmente possuem maior complexidade devido as várias formas e tamanhos que o objeto de interesse pode assumir em qualquer posicão da imagem. Este trabalho possui o objetivo de apresentar os resultados obtidos utilizando as técnicas de janela deslizante e redução da escala da imagem, podemos buscar os objetos de interesse.

II. TRABALHOS CORRELATOS

Entre os vários trabalhos correlatos existentes na área, podemos citar trabalhos na área de detecção de imagens utilizando R-CNN[1], que também faz uso de redes convolucionais profundas e que fornece-se para a arquitetura neural imagens e as localizações dos objetos de interesse, que por sua vez são utilizados para treinar a rede e posteriormente identificar novos objetos em novas imagens.

Pode-se citar trabalho utililizando redes neurais profundas(Deep Learning)[2], ao qual faz uso da mesma

técnica que foi utilizado nesse trabalho, recortando a imagem em pedaços menos e reduzindo a escala da imagem.

III METODOLOGIA

Foi treinado uma rede convolucional fornecendo exemplos positivos(imagens de placas de trânsito) e exemplos negativos(imagens de vários tipos; pessoas, carros, asfalto, arvóres, etc..), cuja intenção era permitir que a rede neural convergisse a ponto de conseguir por meio de suas saídas distinguir as classes de interesse, que é positivo ou negativo, ou seja um placa foi reconhecida ou não foi reconhecida, respectivamente.

Os exemplos fornecidos como positivos foram retirados da base de dados GTSRB[3], que é uma base que contêm mais de 50 mil imagens de placas de vários tipos e que geralmente é uma base de dados utilizado para uma vez detectado a placa em uma imagem, ser possível identificar qual é o tipo da placa. Os exemplos negativos foram retirados da base de dados GTSDB[4], cuja base contêm, diferentemente da GTSRB, imagens de paisagens urbanas, e que na foto podem ou não conter placas de trânsito.

A base de dados GTSRB possui 600 imagens de treinamento e 300 de teste. As 600 imagens pertencentes ao conjunto de treinamento da base GTSDB foram utilizadas como conjunto de exemplos negativos; isso foi feito recortando as imagens em vários fragmentos menores, fornecendo assim os exemplos negativos para o treinamento da rede neural; assim como todas as imagens embora original estão no formato RGB(colorido), foram transformados para o formatos BGR(cinza), para diminuir a complexidade do problema e permitir que a rede alcançasse sua capacidade de generalização de forma mais fácil. Com a finalização do

período de treinamento, as imagens pertencentes a base de dados de teste era utilizado para buscar na imagem a presença de placas de trânsito por meio de duas técnicas, a janela deslizante, ao ocorre um interação por toda a imagem, recortando pequenos pedaços da imagem e inserindo essa imagem na rede neural, que possui o objetivo de reconhecer se naquele fragmento de imagem possui a presença de uma placa de trânsito ou não. Tal processo ocorre, como foi explicado anteriormente em toda a imagem e no final teremos que todas as posições da imagem que foram detectadas placas. Uma segunda técnica aliada ao técnica de janela deslizante, chamada de pyramid, diminui a escala da imagem algumas vezes para que estão em posições diferentes na imagem possam ser detectados.

Para desenvolver a rede convolucional foi utilizado o framework caffe[5], que fornece a possibilidade de construção de redes neurais profundas com pouco esforço, além de possuir diversos arquiteturas de rede neurais disponíveis.

Como escolha da arquitetura foi escolhido o modelo BVLC Reference CaffeNet[6], cuja camada de saída foi alterada para duas classes - positiva e negativa.

IV EXPERIMENTOS

No estágio de experimentação foi utilizado para treinamento tamanhos diferentes das bases de dados e averiguado a evolução da acurácia na detecção das placas, em especial como a taxa de falsos positivos na detecção de placas evolui a medida que o tamanho da base de dados aumenta.

A rede neural foi treinada com os seguintes tamanhos de base de dados: 2 mil, 12 mil, 30 mil e 60 mil imagens para o período de treinamento, estando as classes balanceadas, ou seja, em cada base de dados o número de exemplos positivos é igual ao numero de exemplos negativos. A acurácia obtida durante o ultimo treinamento que utilizou a base de dados com maior tamanho obteve uma acurácia de 82%.

Para utilização dessas bases no framework **caffe** é preciso realizar alguns pré-processamentos nos dados, como a transformação dos dados de treinamento em um banco de

dados lmdb e consequentemente o cálculo das subtrações das imagens pelas suas médias para ajudar na performance durante o treinamento. Com esses passos, é necessário apenas iniciar o treinamento.

V RESULTADOS

A utilização de bases pequenas, como a contendo 2 mil exemplos positivos e negativos mostrou-se suficiente para durante os testes para a detecção das placas - verdadeiros positivos, porém em compensação a quantidade de ocorrência de falsos positivos é grande, o que é de se esperar, visto que com poucos exemplos negativos, a rede neural não recebe informações suficientes para poder aprender a generalizar, e a medida que a base de dados aumenta, a capacidade de verdadeiros positivos mantêm-se estável, porém a taxa de falso de positivos diminui.

As figuras 1 e 2 demonstram um exemplo de comparação entre as bases com 30 e 60 mil exemplos para uma imagem qualquer do conjunto de testes.

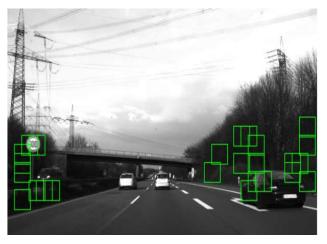


Fig. 1 - Utilização de base de dados com 30 mil exemplos

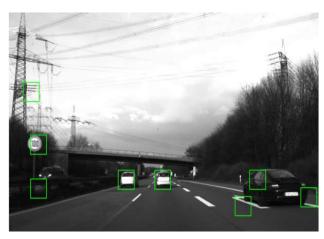


Fig. 2 - Utilização de base de dados com 60 mil exemplos

VI. CONCLUSÃO

Os testes, embora não satisfatórios, pois ainda a ocorrência de muitos falsos positivos, demonstram que com o aumento da base de dados, obtêm-se resultados cada vez melhores.

REFERÊNCIAS

- [1] Girshick, R (2015). Fast R-CNN. page 1
- [2] Szegedy, C., Toshev, A., and Erhan, D.
- [3] Germanic Taffic Sign Benchmark. Disponível em http://benchmark.ini.rub.de/?section=gtsrb&subsection=news.
- [4] Germanic Taffic Sign Benchmark. Disponível em http://benchmark.ini.rub.de/?section=gtsdb&subsection=news
- [5] Caffe Model. Disponível em http://caffe.berkeleyvision.org/model_zoo.html
- [6] Caffe Model. Disponível em http://caffe.berkeleyvision.org/model_zoo.html#bvlc-modellicense