Classificação de Imagens de Sinais de Trânsito Sob Ataques Adversariais

Christian Massao Konishi, Hélio Pedrini

Resumo
O reconhecimento de sinais de trânsito é um campo de pesquisa que tem recebido grande interesse nos últimos anos devido ao seu potencial de uso em várias aplicações, tais como condução autônoma de veículos e assistência ao condutor. Este projeto avaliou o impacto no desempenho de classificadores treinados em bases de dados aumentadas por redes adversárias generativas convolucionais profundas. Para isso, quatro bases de dados diferentes de sinais de trânsito foram testadas. O maior ganho de desempenho observado foi de 98,95% – antes de aumentar a base – para 99,17% – após o aumento da base – na taxa de acurácia.
Palavras chaves:
Sinais de trânsito, aprendizado de máquina, classificação de padrões, visão computacional, análise de imagens

Introdução

Classificar visualmente uma placa de trânsito é uma tarefa que qualquer motorista deve fazer com uma acurácia elevada, visto que o menor dos erros pode provocar uma infração de trânsito com consequências indesejadas ou mesmo um acidente grave. Para automatizar o processo, há uma variedade de abordagens possíveis para análise de imagens, sendo as redes neurais convolucionais o estado da arte neste tópico de pesquisa.

Conseguir um modelo de classificação de sinais de trânsito com elevada eficácia é um objetivo essencial em um contexto de desenvolvimento de sistemas para veículos autônomos, mapeamento das estradas e implementação de tecnologias para auxílio de condução. Um desafio para essas aplicações, contudo, é que cada país possui sua própria legislação de trânsito e, por consequência, suas próprias placas. Dessa forma, a solução não deve restringir os testes a apenas sinais de um banco de dados em específico.

Nesse cenário, este trabalho teve como objetivo analisar o impacto em termos de eficácia do uso de imagens artificialmente criadas por redes adversárias generativas convolucionais profundas (*Deep Convolutional Generative Adversarial Networks* - DCGANs), como forma de balanceamento e aumentação de dados para redes classificadoras.

Bases de Dados

Quatro bases de dados foram utilizadas em nossos experimentos. A primeira foi provida pela The *German Traffic Sign Recognition Benchmark* (GTSRB)[1], contendo 43 diferentes tipos de placas alemães (classes) divididas em um conjunto de treinamento e um conjunto de teste, totalizando mais de cinquenta mil imagens (Figura 2).

Outra base de dados utilizada é a *BelgiumTS Dataset for Classification* (BTSC)[2], contendo 62 diferentes classes de sinais belgas, separadas novamente em um conjunto de treinamento e em um conjunto de teste, somando 7125 imagens (Figura 2).

Outro conjunto de dados utilizado foi a base para classificação de sinais de trânsito croatas (rMASTIF)[3], criada pelo projeto de pesquisa MASTIF, contendo 4044 imagens de treinamento e 1784 de teste (Figura 2).

Por fim, a última base de dados utilizada foi a Tsinghua-Tencent 100K[4]. Trata-se de uma base de sinais de trânsito chineses para detecção e que foi, portanto, adaptada para conter somente imagens de classificação já recortadas. O procedimento consistiu apenas em recortar as imagens de acordo com as anotações previamente fornecidas e em remover as classes com menos de 100 imagens do conjunto de treinamento. A divisão entre teste e treinamento foi feita de acordo com a segregação original da base e resultou, após todo o processo, em 14492 imagens de treinamento e 7234 de teste (Figura 2).

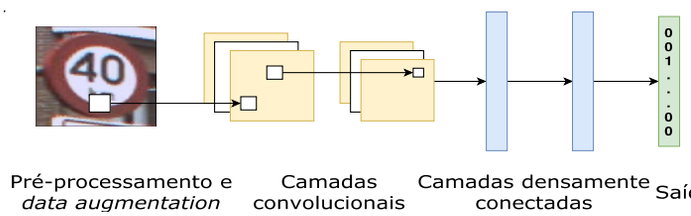

Figura 1: Diagrama da estrutura dos modelos de classificação de imagens, as quantidades de camadas convolucionais e densamente conectadas variam de acordo com a arquitetura.

Após a preparação dos conjuntos de dados, algumas etapas de pré-processamento e aumentação de dados, foram aplicadas. Esses dados então foram fornecidos ao classificador, produzindo uma saída que indica a que classe é mais provável esta amostra pertencer, conforme ilustrado na Figura 1. Para os testes, um classificador robusto e muito conhecido foi utilizado, a *ResNet*-101[5].

1 J. Stallkamp et al., "Man vs. Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition", *Neural Networks*, 2012.

2 Radu Timofte e Luc Van Gool, "Sparse Representation Based Projections", in *British Machine Vision Conference*, 2011, 61.1–61.12.

3 *rMASTIF Traffic Sign Classification Dataset*, [s.d.], http://www.zemris.fer.hr/ kalfa/Datasets/rMASTIF/.

4 Zhe Zhu et al., "Traffic-Sign Detection and Classification in the Wild", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

5 Kaiming He et al., "Deep Residual Learning for Image Recognition", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

(a) GTSRB

(b) BTSC

(c) rMASTIF

(d) Tsinghua-Tencent 100K

Figura 2: Amostras das imagens de sinais de trânsito de cada uma das bases de dados (a-d).

Resultados Experimentais Preliminares

Esta seção tem como objetivo descrever o tratamento das imagens fornecidas para cada base de dados e apresentar os resultados obtidos antes da aumentação de dados através de redes adversárias.

Pré-processamento

Ao longo de testes iniciais, alguns tratamentos nas imagens foram abandonados e, outros, incorporados. Apenas a abordagem final é descrita nesta seção, variando um pouco entre as bases escolhidas.

Para todas as bases de dados trabalhadas, um aumento de nitidez e contraste foi aplicado, utilizando-se um fator de 2 e 1,25, respectivamente. Além disso, as imagens foram recortadas de acordo com anotações fornecidas pelas bases, de forma a remover o máximo de bordas possível. A única exceção é a base belga, na qual as imagens não foram recortadas, tendo em vista que foram detectadas anotações cuja área de interesse está incorretamente delimitada, resultando em imagens inapropriadas.

Aumentação de Dados

Algumas técnicas de aumentação de dados foram aplicadas ao modelo, principalmente para prevenir o sobreajuste das redes que fora observado em momentos iniciais. As alterações feitas foram:

- Rotação: foi aplicada uma rotação em cada imagem em um ângulo aleatório, variando de (-15°, 15°).
- Mudança de perspectiva: foi aplicada uma mudança de perspectiva nas imagens, com um escala de que varia aleatoriamente.
- Ruído Gaussiano: foi aplicado ruído Gaussiano nas imagens com uma variância aleatória entre (0; 0,02) e média 0.

Após o pré-processamento e as técnicas de aumentação de dados, os resultados podem ser observados na Figura 3.

Oversampling e *Undersampling*

Um dos maiores desafios que as bases de dados apresentam é o desbalanceamento na quantidade de imagens por classe. Para contornar este problema, para cada imagem no conjunto foi atribuída uma probabilidade desta figura ser escolhida durante o treinamento, considerando que há reposição após sua utilização. A probabilidade P_i da imagem i ser escolhida, é dada pela normalização da equação abaixo.

$$P'_i = 1/N_i$$

em que N_i é a quantidade de amostras que pertencem à mesma classe da imagem i na base.

Desempenho por Base de Dados

Após realizar o treinamento utilizando o otimizador Adam e técnicas de parada precoce, para cada base de dados, o desempenho do classificador foi calculado. Estes dados obtidos foram compilados na Tabela 2, juntamente aos resultados obtidos ao final dos experimentos, após a aumentação de dados através das DCGANs.

O maior desempenho observado, tanto em acurácia quanto em pontuação F1, foi na base alemã, enquanto o pior resultado foi na base chinesa de sinais de trânsito. Aplicando um tratamento semelhante em todos os casos, os resultados ainda assim foram consideravelmente distintos. Algumas hipóteses podem ser levantadas em relação a essas diferenças.

Flutuações entre diferentes instâncias de treinamento são normais, apesar de insuficientes para explicar a diferença obtida. De fato, é mais provável que as características únicas de cada base tenham um impacto maior na eficácia.

(a) GTSRB

(b) BTSC

(c) rMASTIF

(d) Tsinghua-Tencent 100K

Figura 3: Amostras das imagens das bases de dados (a-d) após o pré-processamento e a aumentação de dados.

Aumentação das bases

Nesta seção, o processo de aumentação das bases de dados através de redes adversárias é apresentado, esclarecendo os hiperparâmetros dos treinamentos e as heurísticas para aumentar as bases de dados, assim como os respectivos resultados obtidos.

A arquitetura das DCGANs utilizadas podem ser observadas na Figura 4, e correspondem aos modelos empregados por Radford et al.[6], com leves alterações no tamanho do vetor de entrada da rede generativa – era 100, originalmente.

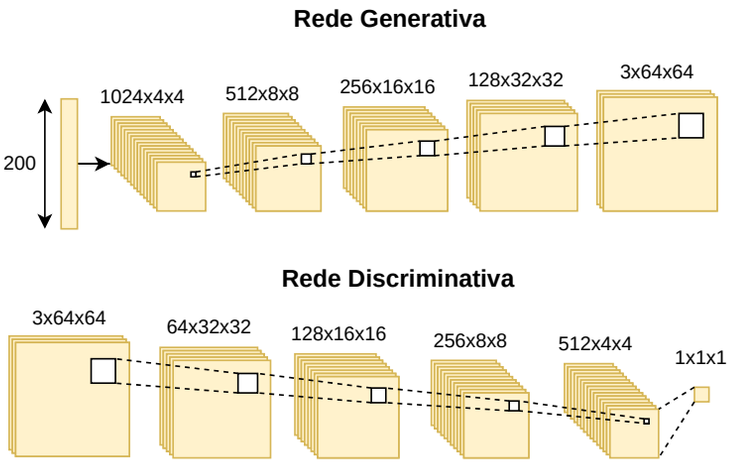

Figura 4: Diagrama das arquiteturas das redes generativa e discriminativa.

O processo para aumentar uma base de dados é custosa, visto que o processo de treinamento é repetido para cada uma das classes da base. Em cada uma das iterações, as redes adversárias utilizam um subconjunto de treinamento, contendo apenas as imagens de uma única classe, sendo os parâmetros aprendidos apropriados apenas para aumentar esta classe de imagem. Por fim, às entradas do discriminador foi adicionado ruído numa distribuição normal padrão e técnicas de *label smoothing* foram empregadas para diminuir a chance de colapso do treinamento.

Figura 5: Amostras das imagens criadas por redes generativas das diferentes bases de dados (a-d).

Os resultados obtidos variaram bastante de acordo principalmente com a quantidade de imagens em cada classe, sendo possível observar casos em que o treinamento das redes adversárias colapsou, quando alimentado por algumas poucas unidades de imagens.

É esperado, portanto, que uma base de dados mais expressiva, como a GTSRB apresente melhores resultados nas imagens sintéticas, principalmente em relação à rMMASTIF e à BTSC, por serem bases menores. Uma amostra aleatória das imagens obtidas após o treinamento foi compilada na Figura 5.

O saldo final de imagens para cada base de dados pode ser conferido na Tabela 1.

Tabela 1: Quantidades de imagens de treinamento na base de dados, antes e depois do processo de aumentação por redes adversárias.

Base de dados	Inicial	Final	Final/Inicial
GTSRB	39.209	125.775	3,21
BTSC	4.557	30.194	6,63
rMASTIF	4.044	15.314	3,79
Tsinghua-Tencent 100K	14.492	80.883	5,58

Resultados Experimentais Finais

Nesta seção, novos testes feitos nas bases de dados, depois de serem aumentadas com redes generativas, foram realizados. Os desempenhos obtidos nesta última etapa serão então comparados com os resultados preliminares.

Treinamento

O treinamento da ResNet-101 foi repetido nas novas bases de dados, utilizando a mesma configuração adotada anteriormente. As mesmas técnicas de pré-processamento e aumentação de dados utilizadas anteriormente foram aplicadas, assim como a taxa de aprendizado e o otimizador adotados foram os mesmos. Duas mudanças, contudo, foram feitas.

As técnicas de *oversampling* e *undersampling* foram descartadas, visto que a aumentação das bases de dados utilizando as redes adversárias equilibrou o desbalanceamento original, removendo o sentido de aplicar essas técnicas. Os hiperparâmetros de paciência e a frequência de validação da técnica de parada precoce foram atualizados para refletir a quantidade mais elevada de imagens processadas por época.

Desempenho por Base de Dados

Novamente, o desempenho dos modelos foi calculado nas bases de dados, mas agora aumentadas. É interessante colocar esses valores ao lado dos obtidos nos resultados preliminares, para verificar que mudanças foram obtidas (Tabela 2).

O melhor cenário observado está na GTSRB, na qual todas as métricas de desempenho aumentaram, seguido pela Tsinghua-Tencent 100K, que também apresentou uma melhoria, e pela rMASTIF e BTSC, cujo desempenho foi reduzido.

É interessante notar que os dois melhores casos ocorreram nas bases com maior quantidade de imagens, o que não é surpreendente, visto que as imagens geradas pelas redes adversárias de maior qualidade estão contidas nelas (Figura 5). Já foi apontado, inclusive, que a quantidade reduzida de imagens pode resultar em um colapso no treinamento.

Tendo isso em mente, a rMASTIF e a BTSC terem desempenho reduzido é esperado, pois suas imagens criadas são de menor qualidade, devido à quantidade muito menor de imagens por classe. O fato do desempenho ser ainda menor

6 Alec Radford, Luke Metz, e Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", *arXiv preprint arXiv:1511.06434*, 2015.

na BTSC provavelmente é devido à quantidade elevada de classes, não acompanhada por um aumento no número total de imagens em relação à rMASTIF, resultando numa quantidade de amostras por tipo de sinal de trânsito mais reduzida.

Por outro lado, em casos não tão extremos de falta de dados, a aumentação de dados através de redes adversárias apresentou resultados promissores, elevando o desempenho, quanto mais ideal for o cenário.

Nesses casos mais apropriados, deve ser levado em consideração, contudo, o consumo de recursos computacionais e humanos para realizar o treinamento de diversas redes adversárias, para cada classe da base de dados. O aumento de desempenho é sólido por se repetir em praticamente todas as métricas de desempenho, entretanto, não foi observada uma mudança muito expressiva e, dependendo do caso, pode não compensar aplicar essa técnica para conseguir ganhos reduzidos.

Conclusões e Trabalhos Futuros

O desempenho de classificadores como a ResNet para o reconhecimento em imagens de sinais de trânsito já é elevado. Este projeto demonstrou que é possível melhorar ainda mais o desempenho de redes classificadoras através de técnicas de aumentação de dados por redes adversárias generativas convolucionais profundas, ainda que com gastos computacionais elevados.

Com base nas dificuldades de treinar essas redes e nos ganhos não tão expressivos, é seguro afirmar que esta técnica de aumentação não deve ser a primeira prioridade em um projeto de classificação, ainda que possa ser levado em consideração em casos específicos, nos quais todo ganho de desempenho, mesmo que pequeno, seja essencial, como é o caso de algumas aplicações da classificação de sinais de trânsito.

Há, neste mesmo tema, outras abordagens interessantes de se testar. O campo das redes adversárias em específico é recente e grandes avanços têm sido feitos desde 2014. Alternativas à DCGAN incluem a Rede Adversária Generativa Wasserstein, cujos resultados podem ser superiores. Expandir os testes para compreender mais arquiteturas de classificadores e mais redes adversárias é uma possibilidade para atingir resultados mais significativos.

Agradecimentos

Os autores são gratos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa de iniciação científica e a todos os doadores e organizadores da Bolsa Alumni pela concessão da bolsa complementar de iniciação científica.

Tabela 2: Comparação do desempenho atingido pela ResNet-101 nas bases de dados avaliadas, antes e depois do processo de aumentação de dados através de redes adversárias.

Base de dados	Acurácia		Precisão		Sensibilidade		Especificidade		Medida F1	
	Antes	Depois	Antes	Depois	Antes	Depois	Antes	Depois	Antes	Depois
GTSRB	98,95%	99,17%	98,21%	98,86%	98,81%	98,85%	99,97%	99,98%	98,45%	98,83%
BTSC	98,29%	95,83%	95,42%	92,42%	96,88%	91,59%	99,97%	99,93%	95,40%	91,00%
rMASTIF	98,65%	98,48%	98,67%	98,46%	97,95%	97,77%	99,96%	99,95%	98,23%	98,05%
Tsinghua-Tencent 100K	97,95%	98,00%	97,37%	98,30%	98,30%	97,80%	99,93%	99,94%	97,78%	98,04%