Aprendizagem Automática Avançada

2022/2023

João Faia – 47051

Tomás Oom – 59447

Classificação de Imagens de Sinais de Trânsito com Recurso a Diferentes Modelos de Aprendizagem Automática

A classificação de imagens é uma área bastante importante da ciência da computação que envolve o desenvolvimento de algoritmos e técnicas para analisar e classificar imagens digitais em diferentes categorias. A importância da classificação de imagens é evidente em diversas áreas, incluindo medicina, indústria, entretenimento e segurança. Uma das aplicações mais úteis é na criação de sistemas de condução autónoma, onde é necessário um modelo, ou um conjunto de modelos, conseguir classificar, objetos, pessoas e a sinalética em redor do veículo. Uma parte importante deste processo é a classificação dos sinais de trânsito, identificando a sua classe e permitindo a tomada de decisão do sistema. Um dataset disponível no Kaggle, "GTSRB - German Traffic Sign Recognition Benchmark", contém um conjunto de imagens de sinais de trânsito alemães que permitem o treino de modelos de classificação.

O objetivo deste projeto é a comparação da performance de três tipos de modelos diferentes para a classificação de imagens de sinais de trânsito fornecidos pelo dataset referido acima. Os modelos a serem testados são multilayer neural networks (MLP), convolutional neural networks (CNN)[2, 3], support vector machines (SVM) e random forests (RF) com diferentes hiperparâmetros. Para cumprir este objetivo, o pré-processamento dos dados é essencial pois cada modelo requer um tipo diferente de input. Por exemplo, no caso de SVM e RF será necessário extrair as features[4] das imagens e ordenar os dados em arrays 2D, enquanto as CNN já recebem como input uma matriz com um formato diferente, por exemplo (50, 32, 32, 3), sendo 50 o número de imagens, 32x32 as dimensões da mesma e 3, o número de channels de cor. Assim, o redimensionamento das imagens antes de se introduzir como input é um passo bastante relevante no pré-processamento dos dados, tendo em consideração que cada modelo tem o seu formato específico. Após este passo, será também necessário normalizar os dados para que os valores dos pixels se enquadrem numa escala de 0 a 1. Isto torna os dados mais consistentes e melhora a convergência dos modelos, considerando que o range dos valores é menor e diminui o enviesamento do modelo. Para alcançar esta normalização, os valores de pixels das imagens têm de ser divididos por 255, valor máximo para uma escala de 8-bits.

Considerando o dataset GTSRB, este é composto por 51839 imagens divididas em 43 classes de sinais de trânsito diferentes. O conjunto de imagens foi dividido num conjunto de treino e de teste com uma partição de 75% e 25%, respetivamente. O set de treino inclui 39209 imagens, sendo necessário, ainda no pré-processamento de dados, perceber a distribuição das classes. Após uma primeira análise (figura 1), é possível verificar um desequilíbrio na distribuição, sendo alguns tipo de imagens muito mais representados que outros. Esta situação pode levar a um enviesamento dos modelos para a classe mais representada, originando classificações incorretas e uma performance mais baixa. Para colmatar este problema, uma fase de data augmentation terá de ser realizada, gerando novas amostras através da aplicação de diversas transformações

às imagens do dataset, como: rotações, espelhagens, mover o objeto da imagem em certa direção, modificar o brilho, contraste ou saturação ou acrescentar ruído gaussiano[5]. Assim, com o conjunto de dados equilibrado podem ser aplicados os modelos de classificação.

No caso da CNN, esta pode ser treinada de origem ou aplicada a técnica de transfer learning, que envolve a utilização de um modelo pré-treinado em datasets de grande volume de imagens como ImageNet[2]. Estes modelos pré-treinados (p.ex: VGG16[6], ResNet[7], Inception[8], etc), permitem acelerar o processo de treino do modelo com os dados do dataset, tendo em conta que a inicialização dos weights já vai ter em conta um conhecimento à priori, sendo apenas necessário realizar o fine-tuning do modelo. Esta vai ser uma opção extra ponderada na fase de treino do modelo, tendo em conta que uma das limitações já consideradas é o baixo poder computacional para treinar um conjunto de dados desta magnitude. Assim, se o poder computacional, de facto, se verificar como uma limitação, o modelo pré-treinado permite que se efetue uma redução no tamanho do conjunto de treino, sem que se perca muita performance.

Por fim, após a análise da performance de todos os modelos, o que obtiver melhor desempenho será escolhido para ser testado com um vídeo de condução numa estrada alemã, onde se irá identificar os sinais em tempo real.

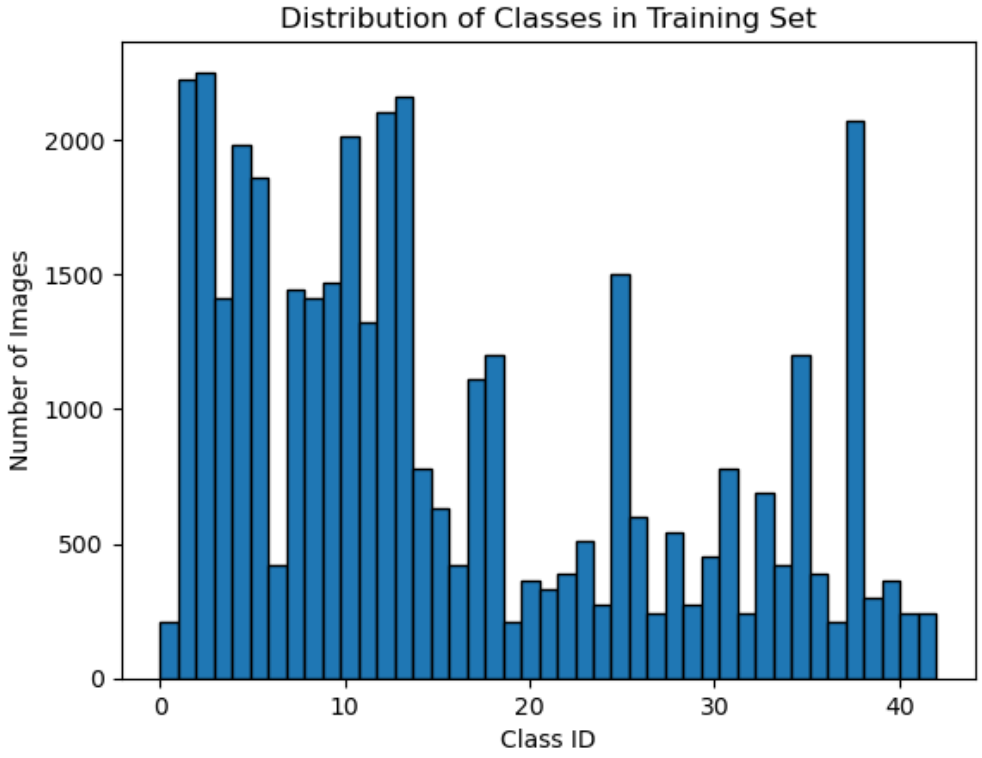

Figura 1 – Representação do desequilíbrio presente nas classes de sinais de trânsito do conjunto de treino, mostrando a necessidade de equilibrar as classes para aumentar a performance dos modelos.

Referências

[1]Ellen, J. S., Graff, C. A., & Ohman, M. D. (2019). Improving plankton image classification using context metadata. Limnology and Oceanography: Methods, 17(8), 439-461.

[2]Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

[3]Bengio, Y., & LeCun, Y. (2007). Scaling learning algorithms towards AI. Large-scale kernel machines, 34(5), 1-41.

[4]Peura, M., & Iivarinen, J. (1997, May). Efficiency of simple shape descriptors. In Proceedings of the third international workshop on visual form (Vol. 5, pp. 443-451).

[5]Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of big data, 6(1), 1-48.

[6]Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[7]He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[8]Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).