Análise da remoção de background no processo de Person Re-Identification

Diógenes Silva
Centro de Informática
Universidade Federal de Pernambuco
Recife, Brasil
dwfs2@cin.ufpe.br

***Resumo*—O processo de Re-Identification, conhecido também pela sigla re-ID, é considerado como um tópico importante em diversas áreas, como: visão computacional e sistemas de vigilância. O objetivo do re-ID é re-identificar pessoas cuja imagem foi capturada por diferentes câmeras localizadas em posições diversas. O re-ID vem ganhando destaque ao longo dos anos e vários estudos vem sendo feitos para melhorar os sistemas atuais. Entre as técnicas para realizar os melhoramentos, tem-se desde o desenvolvimento de métodos de batch normalization até novas estruturas para lidar com os dados (lifelong learning, por exemplo). Neste trabalho, propomos a análise da remoção do background das imagens utilizadas na tarefa de re-ID para verificar o impacto do processo. Infelizmente os métodos usados atualmente possuem um sério problema de overfitting e o background pode ser um fator que influencie nesses resultados, por isso a necessidade da análise. Escolheu-se o dataset Market-1501 e o modelo base para realizar o processo de re-ID foi a ResNet-50, fazendo uso do framework torchreid, uma vez que tem-se facilidade de se trabalhar nele e tanto a implementação quanto os relatórios estão bem descritos, além da documentação disponível descrevendo cada detalhe do método.**

***Index Terms*—Person Re-Identification, Pedestrian Detection, Deep Learning, Image Processing**

I. Introdução

O processo de re-ID busca identificar a mesma pessoa capturada por câmeras em diferentes posições [1]. Um dos tópicos presentes no re-ID é a possibilidade de se localizar criminosos e também pessoas desaparecidas. Sendo assim, o re-ID é uma forma inteligente de se fazer vigilância que vem recebendo uma atenção crescente ao longo dos anos, uma forma de se visualizar este crescimento é pelo número de papers que vem crescendo a cada ano, conforme visto em [2].

Os métodos que obtiveram destaque são baseados em Deep Learning [3-6] e conseguiram avanços significativos através do desenvolvimento desta área da Inteligência Artificial, bem como dos hardwares disponíveis. Tais abordagens tem como objetivo aprender features relacionadas às imagens dos pedestres, de modo que seja possível fazer a re-identificação de câmeras em posições diferentes. Por exemplo, a técnica utilizada neste trabalho [2] gera um vetor representando a pessoa que deve ser comparado com as representações de outros pedestres, em que os vetores mais próximos, em tese, seriam da mesma pessoa.

Diante deste contexto, ainda existem desafios em aberto que não foram investigados de forma intensiva. Entre tais desafios está a influência do background. Geralmente os processos de re-ID se dão através de imagens de pedestres obtidas através de um corte baseado numa bounding box [7,8,9]. Sendo assim, as imagens contém além da imagem do pedestre, o background do local. Esta situação configura um problema, uma vez que tanto a imagem do pedestre quanto a do background são vistas da mesma forma. Nos casos em que os backgrounds de uma mesma pessoa variam entre as câmeras (conforme a Fig. 1), tem-se um viés indesejado que pode atrapalhar no processo de re-ID. Por isso, decidimos analisar essa influência do background nos processos de re-identificação de pessoas.

Figura 1. Imagens do Market-1501 dataset.

As técnicas de Deep Learning tratam as imagens como um todo, sem necessariamente focar em regiões específicas, então duas pessoas muito parecidas mas com a cor do cabelo diferente, poderiam acabar sendo confundidas. Por conta disso, analisar o impacto do background se constitui como uma tarefa relevante acerca da re-ID. Em [10] é possível visualizar o quanto o background influência na re-identificação, validando a necessidade da nossa análise.

No nosso trabalho vamos pela primeira vez investigar o impacto do background utilizando a ResNet-50 [11]. A análise se dá no dataset Market-1501. Para tanto, vamos refazer o dataset visando eliminar o background e uma série de experimentos é realizada. O processo de segmentação é realizado utilizando o detectron2 [12] desenvolvido pelo Facebook. Alguns processamentos na imagem foram necessários, bem como ajustes no processo de segmentação em virtude da baixa resolução das imagens dos pedestres no dataset utilizado.

II. Metodologia

Conforme apresentado em [10], o background influencia na performance no processo de re-ID. Sendo assim, propõe-se uma análise do efeito do background em um método de re-ID específico. Tal método vai ser aplicado em um dataset contendo imagens de pedestres capturadas por câmeras em diferentes posições. Para realizar a análise, o dataset deve ser reconstruído partindo do original através do processo de segmentação para remover o background das imagens. Uma vez com o dataset reconstruído, faz-se o treinamento da CNN e em seguida são feitas as avaliações no conjunto de testes. São definidas duas métricas que vão ser usadas como referência na verificação da performance dos modelos.

Os resultados apresentados em [10], mostram que a remoção do background apresenta melhoras em alguns modelos e piora em outros. Nosso objetivo é verificar o impacto de remover o background utilizando o torchreid e o ResNet-50 como backbone. Opta-se por usar o torchreid [2], uma vez que é um framework bem documentado e também há a possibilidade de manipular facilmente a estrutura da rede, como os hiperparâmetros da CNN. Para realizar a análise, cria-se um novo dataset e em seguida realiza-se a re-ID, sendo escolhido como dataset o Market-1501 [7]. A análise consiste em modificar o Market-1501, removendo o background das imagens e logo após verificar as mudanças nas métricas de mAP e rank quando realizada a re-ID. O processo em si está resumido na Fig. 2.

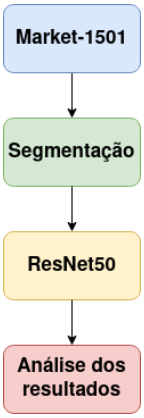

Figura 2. Diagrama referente as etapas da análise.

Quando se fala em remoção de background para re-ID, os trabalhos [10,13] são os que se destacam. Começando por [10], tem-se um trabalho que pode ser considerado como o mais relevante neste quesito, sendo publicado no CVPR 2018, [10] faz uma análise abrangente e traz experimentos que nos instiga a fazer essa investigação acerca do dataset. [10] propõe uma Deep Neural Network baseada em human parsing maps para obter features mais discriminativas, além de realizar data augmentation atribuindo um random background as imagens dos pedestres. Também vale mencionar que os experimentos foram realizados em dois datasets diferentes com imagens com background, sem as pessoas e apenas o background, sem o background e com um background aleatório diferente do original. Tais experimentos são fundamentais, uma vez que mostram o quanto o background influencia nos modelos de re-ID. Já [13] é um trabalho mais recente (2020) e também faz uso da remoção do background, mais especificamente eles propõem duas técnicas, sendo uma delas a supressão gradual do background (GBS, do inglês, gradual background supression). Os resultados foram positivos e foi estabelecida uma estratégia competitiva se comparada com o estado da arte.

III. Implementação

Para fazer a análise, faz-se uso de uma técnica de re-ID, um dataset de pedestres capturados por diferentes câmeras e um elemento segmentador. Os parágrafos seguintes contêm uma breve descrição acerca de cada um dos 3 pontos citados.

O método usado para realizar a re-ID tem como modelo base a ResNet, cuja implementação se dá pela torchreid library. A torchreid é uma library para executar a re-ID usando métodos de Deep Learning e é implementada usando a ferramenta Pytorch, a qual vem ganhando popularidade nos últimos anos e como o nome sugere, trata-se de uma biblioteca na linguagem python. O torchreid permite um rápido desenvolvimento de uma solução end-to-end em que não apenas a rede é treinada, mas a cada 10 epochs (e também após o término do treino) é feita uma avaliação nos dados de teste automaticamente em que são fornecidos o mAP e os ranks. A estrutura geral do torchreid é resumida na Fig. 3.

```
torchreid/
  data/ # data loaders, data augmentation methods, data samplers
  engine/ # training and evaluation pipelines
  losses/ # loss functions
  metrics/ # distance metrics, evaluation metrics
  models # CNN architectures
  optim/ # optimiser and learning rate schedulers
  utils/ # useful tools (also suitable for other PyTorch projects)
```

Figura 3. Estrutura geral do torchreid.

Para realizar o processo de re-ID, fez-se necessário clonar o repositório da torchreid [2] e setar o ambiente de maneira correta, no nosso caso utilizamos o anaconda [14], uma vez que possibilita a flexibilidade de se criar ambientes virtuais, bastando apenas ativar o ambiente para realizar os processos. Com o ambiente montado, criou-se um script chamado de ReId.py. Nele são definidos os parâmetros a serem utilizados. Todo o torchreid foi desenvolvido em python, assim como todos os processos da nossa análise. Utilizamos como modelo a ResNet-50 e os demais detalhes acerca dos parâmetros podem ser verificados no script ReId.py, nele está a configuração estabelecida para o processo de re-ID. Outro ponto importante acerca do processo é a escolha do dataset, nas nossas análises escolhemos o Market-1501.

Vamos agora comentar sobre o dataset. O próprio torchreid já tem uma estrutura pronta para lidar com o dataset escolhido, de tal forma que o seu download é feito de maneira automática. O dataset Market-1501 tem seus detalhes descritos em [7]. Em resumo trata-se de um dataset que contém imagens de pedestres capturadas por 6 câmeras diferentes em um ambiente aberto na Universidade de Tsinghua, na

Figura 4. Imagens de pedestres capturadas por 6 câmeras diferentes.

Fig. 4 tem-se exemplos do dataset ilustrando pedestres desses 6 pontos diferentes. Ao todo são 1501 pessoas no dataset, em que dados de 751 são usados para treino e 750 são usados para teste. Existem 5 arquivos no dataset, que são: bounding_box_test, bounding_box_training, query, gt_bbox, gt_query. Em que as imagens em bounding_box_test, bounding_box_training e query são usadas no processo de re-ID. bounding_box_test contem 19372 imagens para fins de teste, bounding_box_training contém 12936 para fins de treino e query tem 3368 imagens usadas para consulta durante o processo de avaliação da performance do re-ID. O processo de segmentação é então aplicado nas imagens contidas nesses 3 arquivos e o processo de re-ID é treinado novamente e em seguida avaliado.

Figura 5. Exemplo de imagem segmentada utilizando o detectron2.

O processo de segmentação foi realizado através do detectron2 [15]. Trata-se de uma ferramenta que realiza detecção e segmentação no Estado da arte. O detectron2 é uma versão melhorada do detectron, mas não possui um artigo científico ligado ao seu desenvolvimento. Porém, o detectron está atrelado a um trabalho publicado no ICCV 2017 [12]. Embora não se tenha os detalhes do detectron2, falando do detectron, tem-se que o mesmo utiliza uma técnica chamada de mask R-CNN. O mask R-CNN é uma versão extendida do faster R-CNN [16] em que ele adiciona um ramo de segmentação em paralelo aos ramos de classificação e regressão referente aos pontos do bounding box.

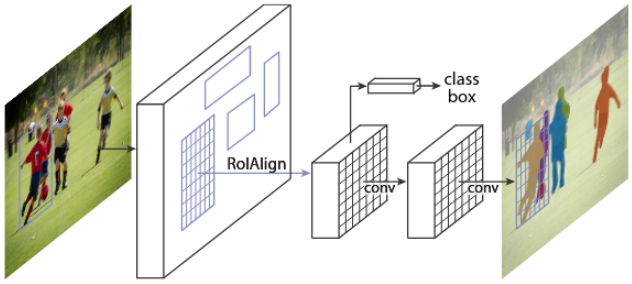

Figura 6. Arquitetura do Mask R-CNN.

Toda a implementação é em python, desde o framework do torchreid até o detectron2, e também a nossa implementação para criar um novo dataset com as imagens do Market-1501 sem o background. O torchreid vai ser usado com o Resnet50 como modelo base e o detectron2 usa o Mask R-CNN. A Fig. 7 ilustra a arquitetura da Resnet50.

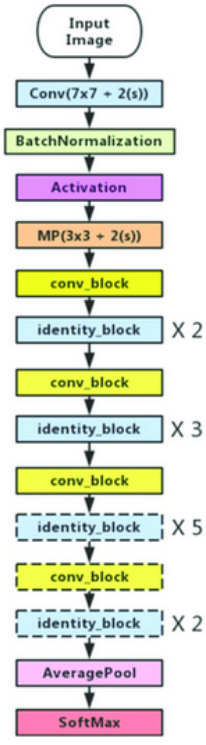

Figura 7. Arquitetura do Resnet50.

IV. EXPERIMENTOS

Para a realização dos experimentos, fez-se a reconstrução do dataset Market1501. Essa reconstrução se deu através do detectron2. O objetivo é de reconstruir o dataset aplicando segmentação e selecionando apenas a região da imagem onde as pessoas estavam contidas. Ao todo foram construídos 2 datasets, um com a segmentação numa qualidade mais baixa, mas em compensação sem pegar nenhum pedaço do background, já o outro continha algumas partes do background, mas a segmentação estava melhor, esses datasets se diferenciam no texto como: sem e com resize, respectivamente. Uma vez reconstruídos os datasets fez-se o re-ID. Conforme a tabela 2, foram feitos 3 experimentos: com o dataset original e com os 2 datasets reconstruídos. Os resultados foram comparados pelas métricas de mAP e rank, que são as métricas comumente usadas nos trabalhos de re-ID.

Dado o dataset Market-1501, vamos aplicar o processo de segmentação em cada imagem e em seguida, usando o ResNet50 como modelo, aplicar o processo de re-ID. No próprio dataset existem as imagens de treino e teste. Em virtude da complexidade da ResNet50 e do tamanho dos dados, os experimentos precisaram ser realizados em um notebook com placa de vídeo GEFORCE GTX 1060 com 6GB.

Devido a baixa resolução das imagens dos pedestres, fez-se necessário realizar alguns ajustes no detectron2 e até nas próprias imagens. A Fig. 8. ilustra os resultados da segmentação, em que os resultados vão melhorando à medida que alguns ajustes são realizados.

Figura 8. Experimentos do processo de segmentação.

Começando pelo detectron2, tem-se um parâmetro de threshold em relação a regiões de interesse, de modo que diminuir o valor vai gerar mais segmentações, porém as chances de ocorrer um falso positivo também aumentam. Contudo, dentre os possíveis elementos a serem segmentados a pessoa é o mais provável, tendo em vista que boa parte da imagem é composta pelo pedestre, sendo assim, o valor do threshold foi reduzido para aumentar a possibilidade de segmentação para localizar o pedestre. Sem esse ajuste algumas imagens mal puderam ser segmentadas devido a baixa resolução da imagem. Porém, ainda não foi o suficiente e o processo ainda estava perdendo parte considerável do pedestre, então decidiu-se aplicar um resize aumentando a altura da imagem e o resultado foi surpreendentemente melhor. Uma vez que a segmentação estava sendo realizada, decidiu-se construir um novo dataset aplicando a segmentação das imagens. Ao todo foram criados 2 datasets, um sem aplicar resize na segmentação e outro aplicando resize. Ambos usando o ResNet50 e os mesmos parâmetros de rede. A diferença entre os dois é a qualidade da segmentação, com resize o resultado é melhor, sem resize a segmentação é feita parcialmente. Um exemplo sem resize é a primeira imagem segmentada da esquerda para direita da Fig. 8. Já a última imagem da Fig. 8 ilustra o resultado de se aplicar o resize.

De posse do novo dataset, aplica-se o re-ID usando o ResNet-50. As métricas utilizadas para avaliar a performance são as métricas utilizadas pelos trabalhos na área de re-ID: mAP (mean average precision) e rank. A mAP se trata da média da AP, a AP é uma métrica famosa em processos de identificação de objetos e que consiste em calcular a precisão e o recall para diferentes valores de threshold. Calcula-se então uma média ponderada da precisão em função da variação do recall, a média aritmética dos valores da AP constitui a mAP. A AP está descrita em (1).

$$AP = \sum_{k=0}^{k=n-1} [Recalls(k) - Recalls(k+1)] * Precisions(k) \tag{1}$$

Sendo (para n = número de thresholds):

$$Recalls(n) = 0, Precisions(n) = 1. \tag{2}$$

Já os ranks funcionam como a acurácia em relação ao número de acertos da classificação dos pedestres, porém existem diferentes ranks, em que o rank-N seria um rank em que a classificação foi correta considerando os N prováveis resultados de classificação.

V. RESULTADOS

Nesta seção são apresentados os resultados nos dois datasets criados. Abaixo segue uma tabela detalhando o dataset utilizado como base.

Tabela 1

Dataset	#IDs (T-G-Q)	#imagens (T-G-Q)
Market1501	751-750-751	12936-3368-15913

São aplicadas as imagens do Market1501 a segmentação do detectron2 e assim gera-se um novo dataset. Aqui dividimos em com e sem resize, porque esta operação influenciava na detecção que estava com problemas devido a baixa resolução das imagens. Abaixo seguem os resultados:

Tabela 2

Metodo	Backbone	Dataset	mAP	Rank-1	Rank-5	Rank-10
Torchreid	ResNet50	Market1501	66.8	84.5	93.8	96.1
Torchreid	ResNet50	Nosso (sem resize)	49.5	70.5	85.7	89.9
Torchreid	ResNet50	Nosso (com resize)	60.7	79.8	91.3	94.4

A proposta é fazer uma análise do quanto o re-ID poderia variar eliminando o background. Pelos resultados apresentados, tem-se que uma segmentação parcial em algumas imagens, como o do caso do dataset sem resize, apresentou um

resultado significativamente pior. Já com o resize o resultado é bem melhor, mas ainda assim fica abaixo do re-ID com o background das imagens. A Fig. 9 mostra uma comparação do mAP para as epochs da avaliação nos 2 datasets gerados.

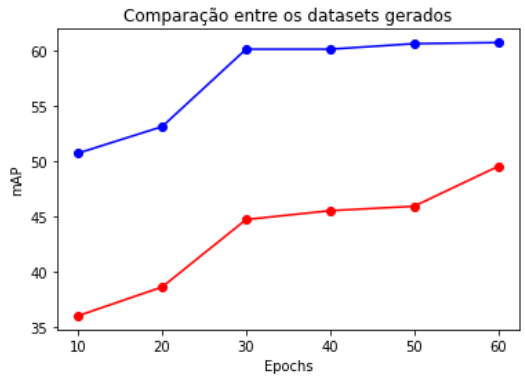

Figura 9. Experimentos do processo de segmentação. Em azul o dataset com resize e em vermelho sem resize.

Fica nítido que o re-ID funciona melhor quanto mais informação temos da imagem original. Em relação ao dataset sem resize a análise é mais simples, porque segmentar apenas parcialmente acaba realmente sendo pior, mas com o resize a segmentação funciona melhor, mas o resultado foi pior do que com o background. Como com o resize tem-se o pedestre como um todo, a pessoa fica em evidência na imagem e esperava-se no mínimo o mesmo resultado do que quando havia o background, porém não foi o que ocorreu.

VI. Conclusão

A área de re-ID vem ganhando relevância nos últimos anos, mas mesmo assim existem desafios em aberto. Por conta disso, nós decidimos fazer análises envolvendo o background da imagem para verificar o seu impacto no re-ID e se de fato poderia haver melhorias para o re-ID, uma vez que o background varia dependendo de qual câmera a imagem foi capturada. Para tanto, fizemos uso de um framework próprio para re-ID utilizando a ResNet50 como backbone.

No final não foi possível observar melhoras no re-ID, em que o dataset sem resize, cuja a segmentação foi inferior ao do outro dataset com resize, apresentou uma piora significativa. Já o dataset com resize foi bem melhor do que o sem resize, mas ainda foi inferior ao dataset original.

Como trabalhos futuros existem as possibilidades de se fazer um segmentador especificamente para problemas de re-ID, uma vez que as imagens possuem baixa resolução e mesmo os segmentadores Estado da Arte como o detectron2 não foram o suficiente, embora funcionem bem em imagens de maior resolução. Utilizou-se apenas um backbone que foi o ResNet50, então ainda existe a possibilidade de se usar outros backbones ou criar uma rede própria, uma sugestão é observar cuidadosamente a CNN usada no trabalho que faz remoção de background [10] e usá-la como guia para aplicar as melhorias encontradas nos métodos de Deep Learning Estado da arte. Neste caso fez-se a análise apenas no Market1501, isto é, ainda existem mais datasets a serem trabalhados para realizar uma análise mais profunda e precisa acerca do impacto do background nas técnicas re-ID.

Referências

[1] Marco Cristani, Vittorio Murino, Chapter 10 - Person re-identification, Editor(s): Rama Chellappa, Sergios Theodoridis, Academic Press Library in Signal Processing, Volume 6, Academic Press, 2018, Pages 365-394, ISBN 9780128118894, https://doi.org/10.1016/B978-0-12-811889-4.00010-5.

[2] ZHOU, Kaiyang; XIANG, Tao. Torchreid: A library for deep learning person re-identification in pytorch. arXiv preprint arXiv:1910.10093, 2019.

[3] CHOI, Seokeon et al. Meta Batch-Instance Normalization for Generalizable Person Re-Identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. p. 3425-3435.

[4] ZHOU, Kaiyang et al. Learning generalisable omni-scale representations for person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[5] SOVRASOV, Vladislav; SIDNEV, Dmitry. Building Computationally Efficient and Well-Generalizing Person Re-Identification Models with Metric Learning. In: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021. p. 639-646.

[6] WU, Guile; GONG, Shaogang. Generalising without Forgetting for Lifelong Person Re-Identification. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2021. p. 2889-2897.

[7] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, Q. (2015). Scalable person re-identification: A benchmark. In ICCV.

[8] Li, W., Zhao, R., Xiao, T., and Wang, X. (2014). Deepreid: Deep filter pairing neural network for person re-identification. In CVPR.

[9] Ristani, E., Solera, F., Zou, R., Cucchiara, R., and Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. In ECCVW.

[10] TIAN, Maoqing et al. Eliminating background-bias for robust person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 5794-5803.

[11] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In CVPR.

[12] HE, Kaiming et al. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2961-2969.

[13] TANG, Yingzhi et al. Person re-identification with feature pyramid optimization and gradual background suppression. Neural Networks, v. 124, p. 223-232, 2020.

[14] Anon, 2020. Anaconda Software Distribution, Anaconda Inc. Available at: https://docs.anaconda.com/.

[15] Yuxin Wu and Alexander Kirillov and Francisco Massa and Wan-Yen Lo and Ross Girshick , Detectron2 (2019), https://github.com/facebookresearch/detectron2.

[16] GIRSHICK, Ross. Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2015. p. 1440-1448.