Augusto Maillo Queiroga de Figueiredo

Re-Identificação de motociclistas em cenários de vigilância

Belo Horizonte, Minas Gerais
2023

Augusto Maillo Queiroga de Figueiredo

Re-Identificação de motociclistas em cenários de vigilância

Relatório final para o Projeto Orientado em Computação do curso de Bacharelado em Ciência da Computação na Universidade Federal de Minas Gerais. Relatório desenvolvido pelo aluno Augusto Maillo Queiroga de Figueiredo, orientado pelo Prof. William Robson Schwartz.

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Orientador: William Robson Schwartz

Belo Horizonte, Minas Gerais
2023

Sumário

1 Introdução

Uma vez que motocicletas são muito presentes no trânsito e, em virtude de sua agilidade, são frequentemente utilizadas para atividades criminosas, o desenvolvimento de tecnologias capazes de apoiar a vigilância de motociclistas se faz necessário.

Embora o campo de Re-Identificação de pessoas já seja bem estabelecido, existem diferenças fundamentais nos cenários esperados para esta tarefa e os cenários de vigilância de motociclistas. Sendo assim, a tarefa de Re-Identificação de motociclistas acaba sendo um campo apartado, possuindo características e desafios próprios.

Atualmente existem dois conjuntos de dados amplamente utilizados para a tarefa de re-identificação de motociclistas: o MoRe [Figueiredo et al. 2021] e o BPReid [Yuan, Zhang e Wang 2018]. Embora sejam conjuntos de dados distintos, ambos fornecem dados relevantes para a análise de desempenho de algoritmos de re-identificação nesse contexto. Apesar da existência de um baseline forte para esses conjuntos, ainda há espaço para melhorias, visando alcançar uma solução mais precisa e eficiente para a tarefa. Nesse sentido, a metodologia proposta neste trabalho busca superar esses baselines, utilizando técnicas avançadas de aprendizado de máquina e um pipeline de treinamento projetado para lidar com as particularidades desse tipo de cenário de vigilância.

Ao final da primeira parte do projeto, foi possível atingir uma nova versão da implementação do baseline de Re-Identificação do conjunto MoRe mais performática e simples de realizar testes. Com isso, os experimentos que serão realizados nesta etapa poderão ser realizados com maior agilidade e em maior número.

1.1 Objetivos Gerais

Este trabalho visa atualizar a implementação do *baseline* atual do MoRe para facilitar a experimentação com o conjunto de dados. Junto a isso, espera-se ao final fornecer uma metodologia para a tarefa de Re-Identificação de motociclistas em cenários reais de vigilância, apresentando um modelo e um *pipeline* para treino, que supere o *baseline* atual.

2 Referencial Teórico

A tarefa de re-identificação consiste no processo de corresponder imagens de uma pessoa em diferentes câmeras ou em diferentes momentos, o que a faz ter grande importante em cenários de vigilâncias. Contudo, é uma tarefa particularmente desafiadora quando tratamos de motociclistas, devido às suas características como capacetes, vestimentas e diferentes posturas no guidão, que dificultam a identificação precisa.

Essa é uma tarefa complexa que requer a aplicação de técnicas de aprendizado de máquina para reconhecer os mesmos indivíduos em diferentes câmeras ou momentos. Atualmente na literatura científica existem diversas abordagens para a Re-Identificação de pessoas com base em características visuais, as quais utilizam técnicas de aprendizado de máquina profundo para extrair representações vetoriais das imagens.

Apesar da importância desta tarefa, existem apenas 2 conjuntos de dados focados em motociclistas e em cenários de vigilância, o *dataset* BPReid [Yuan, Zhang e Wang 2018] e o *dataset* MoRe [Figueiredo et al. 2021]. Embora ambos os conjuntos incluam imagens de motociclistas, a natureza das imagens que compõe estes conjuntos são fundamentalmente distintas. Enquanto o BPReid é composto por imagens capturadas em um campus universitário em cenário controlado, as imagens que compõe o MoRe são provenientes de câmeras de vigilância reais, trazendo maior desafio para o conjunto.

Ainda que a tarefa de re-identificação de motociclistas seja distinta da re-identificação de pessoas, esta última servirá como base para as metodologias e modelos que serão avaliados neste trabalho, buscando entender como podemos aproveitá-las neste novo cenário. De forma geral, trabalhos de re-identificação estudam o papel essencial do conceito de *metric learning* na otimização do desempenho dos modelos. Esta é uma abordagem no campo de aprendizado de máquina que se concentra na aprendizagem de representações semânticas significativas para os dados. No contexto da re-identificação de motociclistas, a aplicação eficaz desse tipo de técnica pode superar desafios específicos, como variações nas posturas de condução, presença de capacetes e diferenças nas vestimentas. O objetivo de forma geral é aproximar, em um espaço de alta dimensão, as representações semânticas dos mesmos indivíduos e distâncias aquelas de sujeitos diferentes [Wang et al. 2017, Oord, Li e Vinyals 2019].

3 Experimentos

Nesta seção, serão descritas as técnicas e experimentos que foram realizados neste trabalho.

Todos os experimentos foram realizados utilizando a divisão entre treino e teste proposta em [Figueiredo et al. 2021], onde metade dos indivíduos são utilizados para treino e metade para teste. Além disso, foram mantidas as estratégias em relação ao número de épocas e protocolo de decaimento para a taxa de aprendizado.

3.1 Construção de lotes de treinamento

Em problemas de re-identificação, os modelos são treinados para reconhecer objetos em diferentes contextos e condições, apresentando desafios singulares em comparação com tarefas convencionais de classificação de imagens. A construção cuidadosa de lotes de treinamento desempenha um papel crucial na eficácia desse processo. Ao contrário de tarefas comuns de classificação, onde a presença isolada de uma instância de um objeto é suficiente, a re-identificação demanda uma consideração mais profunda das características específicas que distinguem objetos semelhantes. Ao criar lotes de treinamento, é fundamental incorporar uma diversidade representativa de objetos e suas variações contextuais, como, por exemplo, diferentes ângulos de visão, condições de iluminação e fundos. Esta abordagem proporciona ao modelo uma compreensão mais abrangente das nuances visuais associadas aos objetos, capacitando-o a realizar uma re-identificação mais precisa em condições diversas.

Visto a criticidade desse ponto, uma das hipóteses levantadas neste trabalho é de que utilizar técnicas de *hard sampling* trariam ganhos expressivos em relação ao *baseline* aqui adotado. Essa abordagem não é nova e resultados expressivos já foram reportados em diferentes contextos [Chen et al. 2020, Sheng et al. 2020].

Especificamente para o escopo deste trabalho, foi explorada a estratégia de *hard sampling* na construção de lotes para a função de perda Quadruplet Loss[Chen et al. 2017]. Nessa abordagem, ao fixar uma âncora, a primeira etapa consiste em selecionar, de maneira aleatória, uma das duas amostras positivas mais distantes da âncora no espaço vetorial. Em seguida, como primeira amostra negativa, escolhe-se aleatoriamente uma amostra de um sujeito diferente entre os N mais próximos da âncora. Por fim, é selecionada também aleatoriamente uma amostra de um terceiro sujeito entre os N mais próximos da primeira amostra negativa. A intenção é fornecer exemplos desafiadores para o modelo se aproximar quando se trata do mesmo sujeito e se distanciar quando não se trata do mesmo. Essa abordagem visa enriquecer o treinamento ao expor o modelo a casos mais complexos e representativos.

Contudo, um problema surge com essa estratégia. Como o modelo está sendo atualizado a cada passo do treinamento, a representação que gera das imagens também é alterada. Porém, recomputar em todo passo do treinamento essas representações seria inviável, pois levaria a um aumento expressivo do tempo de treinamento. Para solucionar isso, a implementação feita dessa técnica reaproveita todas as predições feitas pelo modelo durante uma época de treino. Ao final de cada época as distâncias entre os vetores-representação é recalculada e utilizada em toda a próxima época na geração dos lotes. Com essa abordagem, incluímos apenas o tempo de atualização das representações de cada amostra e o tempo de recomputar a matriz de distâncias, o que, na prática, significou um tempo inexpressivo frente ao treinamento completo.

Os experimentos realizados nesse escopo buscaram avaliar como o parâmetro N dessa metodologia impacta no resultado, bem como os pesos dados para as funções de perda.

3.2 Novas funções de perda

A escolha e aplicação adequada de funções de perda desempenham um papel fundamental na resolução de problemas de re-identificação, conferindo direção e coerência ao treinamento dos modelos. Em particular, as funções de perda têm a responsabilidade de quantificar a discrepância entre as representações aprendidas para instâncias positivas e negativas, desempenhando um papel central na capacidade do modelo de discernir entre objetos similares. No contexto aqui tratado, onde a precisão na diferenciação de objetos é o ponto-chave, as funções de perda desempenham um papel crucial ao, por exemplo, estabelecer uma margem entre exemplos positivos e negativos. Existem diferentes estratégias inovadoras que buscam aprimorar a distinção entre instâncias, visando uma aprendizagem mais robusta e eficaz. Dessa forma, a seleção criteriosa e a configuração adequada dessas funções são elementos fundamentais na elaboração de uma abordagem bem-sucedida.

Nestre trabalho, foram selecionadas duas funções de perda distintas, a Angular Loss e a InfoNCE Loss, ambas com propósitos específicos na melhoria da re-identificação de objetos. A hipótese é de que a inclusão dessas funções no modelo baseline, em conjunto com a *Quadruplet Loss*, traria resultados significantes.

A Angular Loss [Wang et al. 2017] é uma função que visa minimizar os ângulos entre os vetores representativos de instâncias positivas, pertencentes ao mesmo sujeito. A lógica por trás dessa escolha reside na natureza da distância de cossenos utilizada durante a inferência. Ao reduzir os ângulos entre vetores representativos de um mesmo sujeito, a Angular Loss busca otimizar a relação de similaridade, alinhando-se com a métrica de cosseno empregada para avaliar a proximidade entre representações durante o processo de re-identificação.

Por outro lado, a InfoNCE Loss [Oord, Li e Vinyals 2019], "Noise Contrastive Esti-

mation Loss,"baseia-se na ideia de contraste entre instâncias positivas e negativas para aprender representações robustas. Essa abordagem busca maximizar a similaridade entre exemplos positivos e minimizar entre exemplos negativos, estimulando o modelo a focar em características discriminativas para a tarefa em questão. A escolha da InfoNCE Loss foi motivada pelo seu sucesso em outras tarefas e na sua capacidade de impulsionar a aprendizagem de representações úteis e distintas, o que se mostra crucial na re-identificação de objetos.

Estas funções foram avaliadas já utilizando a abordagem de construção de lotes de treinamento descrita na seção 3.1. Foram feitos experimentos no sentido de verificar quais combinações de pesos para as funções de perda trariam os melhores resultados.

3.3 Novo *backbone*

As redes backbones, no contexto da geração de vetores representativos para sujeitos, desempenham um papel central na eficácia dos modelos de re-identificação. Estas redes constituem a espinha dorsal do sistema, sendo responsáveis por extrair características relevantes das imagens de entrada e transformá-las em representações semânticas compactas e discriminativas (vetores em um espaço de alta dimensão). A escolha apropriada do *backbone* é essencial para a capacidade do modelo de capturar com precisão as nuances visuais que distinguem sujeitos em ambientes diversos. A eficácia dessas redes é determinante não apenas na qualidade das representações vetoriais geradas, mas também na capacidade do modelo de generalizar e re-identificar sujeitos em diferentes condições.

Foi escolhido neste trabalho a EfficientNet [Tan e Le 2020] como backbone para avaliação, escolha essa orientada pela necessidade de equilibrar a eficiência computacional com a capacidade de capturar informações discriminativas. A EfficientNet destaca-se por seu design eficiente, que utiliza um mecanismo de escala composta para otimizar simultaneamente a largura, a profundidade e a resolução da rede. Sua arquitetura busca atingir um equilíbrio ideal entre desempenho e eficiência, resultando em um modelo mais leve em comparação com redes tradicionais, como a *ResNet* ou a *VGG*.

A hipótese subjacente à escolha da *EfficientNet* é que, mesmo sendo uma rede mais leve, ela mantém a capacidade de extrair características discriminativas essenciais para a tarefa de re-identificação de sujeitos. A eficiência dessa arquitetura sugere a possibilidade de alcançar um desempenho competitivo em cenários de restrição computacional, sem comprometer a qualidade das representações vetoriais geradas. Ao explorar a *EfficientNet* como backbone, busca-se avaliar a viabilidade de empregar redes mais leves sem sacrificar a precisão na re-identificação, apresentando uma abordagem promissora para aplicações práticas em ambientes de recursos limitados.

4 Resultados

Nesta seção serão apresentados os resultados dos experimentos descritos na seção 3. As métricas aqui reportadas são a *r1*, taxa de reconhecimento no rank 1, e *mAP*, *mean average precision*, que computa a média da precisão média com base na curva de *precision* x *recall*. Os experimentos foram executados utilizando um processador Intel Core i3-10100f e uma GPU NVIDIA GTX 1080.

4.1 Avaliação da estratégia de amostragem

A tabela 1 mostra os resultados dos experimentos conduzidos no sentido de avaliar a estratégia de *hard sampling*. Além de comparar diferentes valores para N, parâmetro que define quantas amostras difíceis avaliar antes de sortear uma, foram avaliados diferentes combinações de pesos para as funções de perda. Estes resultados mostram como o modelo é sensível a estas combinações. Em especial, nos dá um norte sobre como definir o alvo principal entre as funções de perda. Ainda assim, é possível ver o expressivo ganho trazido por essa abordagem. A métrica de *r1* apresentou um ganho de 11 p.p., o que demonstra o sucesso dessa técnica em nosso contexto.

Método	**Parâmetros**	**Pesos das funções de perda**			**Métricas**	
	N	Cross-entropy	Quadruplet	Center	r1	mAP
	10	0.2	1,0	10^{-2}	**73.63**	**78.59**
	10	0.1	1.0	10^{-3}	72.10	77.08
Hard-sampling	10	0.1	1.0	10^{-1}	66.74	72.25
	50	1.0	1.0	10^{-3}	68.65	73.94
	50	1.0	1.0	10^{-2}	72.21	77.01
Baseline	-		-		62.53	68.53

Tabela 1 – Experimentos avaliando a técnica de Hard-Sampling

4.2 Avaliação das funções de perda

Já a tabela 2 traz os resultados dos experimentos que avaliam a inclusão de novas funções de perda. Em especial foram experimentas diferentes combinações de pesos para as funções. Um ponto importante a se ressaltar aqui é que os resultados reportados representam os experimentos onde foi possível fazer com o que o modelo convergisse seu aprendizado. Muitas outras combinações possíveis para as funções não conseguiram atingir um ponto onde o modelo pudesse avançar com o seu aprendizado. Como mencionado anteriormente, com exceção do modelo baseline, todos os experimentos aqui já utilizam a ideia de *hard*

sampling. Com isso, embora o resultado tenha apresentado uma singela melhora quando comparado ao apresentado na tabela 2, não é uma diferença tão significativa e pode ser atribuída ao acaso, visto que o modelo e o treinamento estão envolvidos em diversas gerações de números aleatórios.

Função de perda	**Pesos das funções de perda**				**Métricas**	
	Cross-entropy	Quadruplet	Center	Loss testada	r1	mAP
Angular Loss	1.0	1.0	10^{-2}	1.0	73.60	77.99
	1.0	1.0	10^{-2}	0.5	66.52	68.12
	1.0	1.0	10^{-2}	0.1	55.21	59.83
InfoNCE Loss	1.0	2.0	10^{-3}	0.5	73.93	78.67
	1.0	5.0	10^{-2}	0.5	**74.15**	**78.94**
	1.0	5.0	10^{-2}	0.5	73.74	78.61
Baseline	-				62.53	68.53

Tabela 2 – Experimentos avaliando diferentes funções de loss

4.3 Avaliação da *EfficientNet* como *backbone*

Por fim, os últimos experimentos feitos avaliaram como um diferente *backbone*, aqui no caso uma rede com um número menor de parâmetros que a *ResNet50* e consequentemente mais leve, representada pela *EfficientNet*, conseguiria performar nesse conjunto de dados.

Contudo, os primeiros experimentos já mostraram o grande desafio dessa hipótese. A tabela 3 apresenta o melhor modelo obtido com a *EfficientNet*, que é muito aquém do baseline. Diversos experimentos foram realizados, alterando diversos hiper-parâmetros como taxa de aprendizado, margens da Quadruplet Loss, etc. Porém, a maioria dos modelos não apresentou convergência desde as primeiras épocas ou não conseguiu prosseguir com o aprendizado a partir da metade do processo. A hipótese para esse comportamento é de que a menor capacidade da rede, pilar de sua maior eficiência computacional, não é suficiente para discriminar os sujeitos. A representação das imagens que, na ResNet50, possuía 2048 dimensões, nesse caso tem 1280 dimensões, o que já mostra uma redução drástica. Embora de fato essa abordagem tenha apresentado um tempo de treinamento e inferência bem menor, ela não se justifica, visto que demonstra uma assertividade bem menor.

Modelo	**Métricas**	
	r1	mAP
EfficientNet como *backbone*	54.47	60.32
Baseline	**62.53**	**68.53**

Tabela 3 – Experimento avaliando a EfficientNet

4.4 Comparação dos resultados

Por fim, a imagem 1 traz uma comparação entre as melhores versões dos modelos testados. Aqui fica bem evidente como o principal ganho obtido vêm da estratégia de amostragem apresentada. Embora a Angular Loss e a InfoNCE Loss tenham sido capazes de apresentar resultados levemente melhores, novamente, são ganhos tímidos que podem ser atribuídos ao acaso, visto a abundância de processos aleatórios envolvidos. Além disso, é importante destacar também como a *EfficientNet* não é capaz de trazer os resultados esperados. Junto a grande dificuldade de conseguir com que o modelo iniciasse o processo de convergência, seu melhor resultado foi muito abaixo do já alcançado pelo baseline.

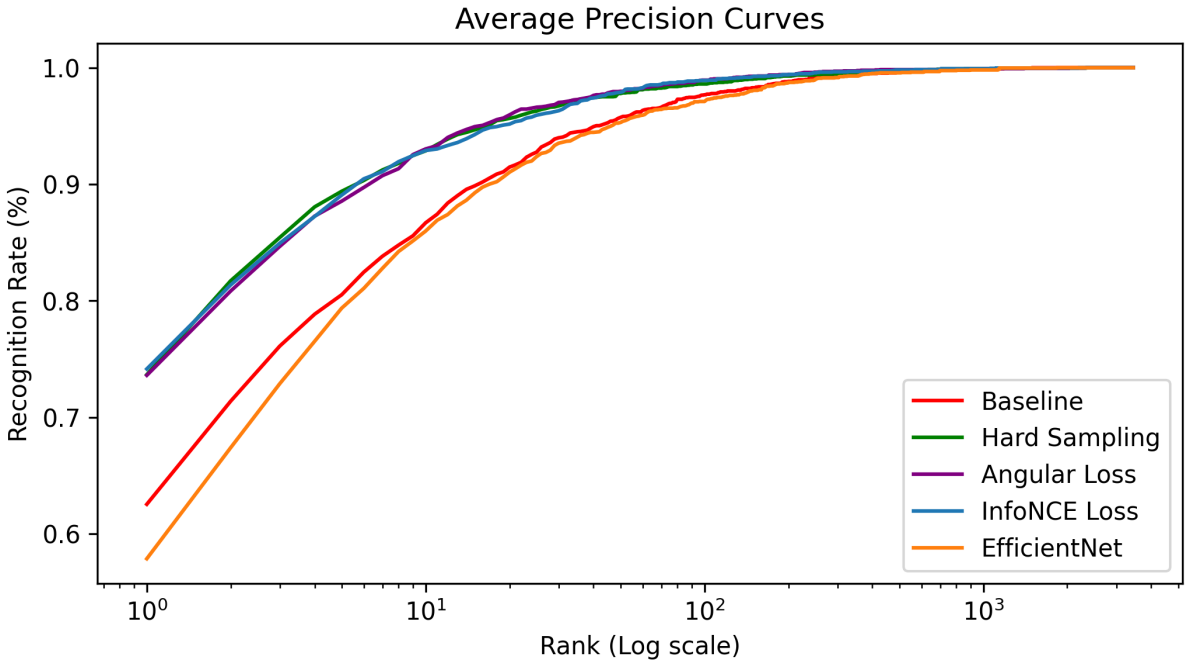

Figura 1 – Comparação entre os melhores modelos de cada experimento

5 Conclusão

Em conclusão, este estudo abordou estratégias e técnicas fundamentais para aprimorar a re-identificação de sujeitos em ambientes desafiadores. A estratégia de amostragem adotada, conhecida como *hard sampling*, mostrou-se crucial para a melhoria do aprendizado do modelo, confirmando nossa primeira hipótese ao proporcionar exemplos mais desafiadores e direcionando eficientemente sua capacidade de discriminação entre sujeitos. A implementação dessa abordagem não apenas se revelou eficaz, mas também demonstrou ser computacionalmente viável, não impondo um ônus significativo no tempo de treinamento.

Os experimentos com diferentes funções de perda revelaram *insights* valiosos sobre a importância da escolha adequada dessas funções na modelagem. A Angular Loss, ao incorporar a noção de ângulo entre vetores, e a InfoNCE Loss, com sua abordagem contrastiva, forneceram perspectivas distintas, mas a estratégia de *hard sampling* destacou-se como o componente mais impactante na melhoria do desempenho. A hipótese de que essas funções de perda trariam ganhos relevantes não se mostrou verdadeira, visto que os ganhos apresentados se revelaram pouco significativos.

A avaliação da *EfficientNet* como *backbone* revelou que, apesar de sua eficiência computacional notável, a rede enfrentou dificuldades em aprender representações discriminativas suficientes para a tarefa de re-identificação. A análise desses resultados sugere que, em alguns casos, a eficiência computacional pode ser comprometida pela capacidade reduzida da rede em aprender nuances complexas do problema em questão.

Em resumo, este trabalho ressalta a importância do componente de *metric learning*, da construção inteligente de lotes de treinamento e da escolha cuidadosa de funções de perda na resolução de desafios inerentes à re-identificação de sujeitos. O objetivo inicial de propor uma estratégia de modelagem que superasse o *baseline* para o dataset More foi totalmente atingido, embora nem todas as hipóteses tenham se mostrado verdadeiras. Junto a isso, as conclusões obtidas fornecem percepções valiosas para futuras pesquisas nesse campo, destacando a necessidade de considerar tanto estratégias inovadoras quanto a eficiência computacional ao projetar abordagens para tarefas similares.

Referências

[Chen et al. 2020]CHEN, K. et al. Hard sample mining makes person re-identification more efficient and accurate. *Neurocomputing*, v. 382, p. 259–267, 2020. ISSN 0925-2312. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231219316984>.

[Chen et al. 2017]CHEN, W. et al. Beyond triplet loss: A deep quadruplet network for person re-identification. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, 2017. p. 1320–1329. ISSN 1063-6919. Disponível em: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.145>.

[Figueiredo et al. 2021]FIGUEIREDO, A. et al. More: A large-scale motorcycle re-identification dataset. In: *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.: s.n.], 2021. p. 4033–4042.

[He et al. 2016]HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778.

[Oord, Li e Vinyals 2019]OORD, A. van den; LI, Y.; VINYALS, O. *Representation Learning with Contrastive Predictive Coding*. 2019.

[Qi e Su 2017]QI, C.; SU, F. Contrastive-center loss for deep neural networks. In: *2017 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2017. p. 2851–2855.

[Sheng et al. 2020]SHENG, H. et al. Mining hard samples globally and efficiently for person reidentification. *IEEE Internet of Things Journal*, v. 7, n. 10, p. 9611–9622, 2020.

[Tan e Le 2020]TAN, M.; LE, Q. V. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. 2020.

[Wang et al. 2017]WANG, J. et al. Deep metric learning with angular loss. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017.

[Yuan, Zhang e Wang 2018]YUAN, Y.; ZHANG, J.; WANG, Q. Bike-person re-identification: A benchmark and a comprehensive evaluation. *IEEE Access*, PP, p. 1–1, 09 2018.