

UTILIZAÇÃO DE TÉCNICAS DE DEEP LEARNING PARA REIDENTIFICAÇÃO DE PESSOAS

PABLO FERNANDES DIAS*, LUIZ ALBERTO PINTO*, DANIEL CRUZ CAVALIERI*, FLÁVIO GARCIA PEREIRA*

**Instituto Federal do Espírito Santo
Rodovia ES-010 – Km 6,5 - Manguinhos
29173-087 – Serra – Es*

Emails: pablo.vix@bol.com.br, luiz.pt@ifes.edu.br, daniel.cavaliere@ifes.edu.br, flavio.garcia@ifes.edu.br

Abstract— The person reidentification problem (Re-ID) consists in recognizing a person of interest who had previously appeared in images taken by distinct cameras. This issue has been appealing due matters of security and the evolution of surveillance and monitoring networks. However, many challenges are yet to be overcome. This work explores the problem based on deep learning techniques. The performance of convolutional neural networks was investigated under different instances of the problem, and the effects on the classification of the use of various preprocessing techniques, varying epochs during the training phase, number of classes considered, and also, the effect of image augmentation. The results show that even with the intrinsic difficulties of the problem, performance improvements can still be obtained under proper configuration conditions.

Keywords— Reidentification, Deep Learning, Transfer Learning, Image Augmentation

Resumo— O problema de reidentificação consiste em reconhecer indivíduos de interesse, que apareceram em imagens de uma câmera distinta em algum momento anterior. Este tema tem se mostrado relevante devido a questões de segurança e a evolução das redes de monitoração e vigilância. Porém, muitos desafios ainda estão por ser superados. O presente trabalho explora o problema através de abordagens baseadas em técnicas de deep learning. Foi investigados o desempenho de Redes Neurais Convolucionais sob diferentes instâncias do problema, e foi estudado o efeito sobre a classificação da utilização de diversas técnicas de pré-processamento, variação do número de épocas na fase de treino, número de classes consideradas, e principalmente, o efeito da aumentação de imagens. Os resultados mostram que, mesmo com as dificuldades inerentes ao problema, sob condições de configurações adequadas melhores desempenhos podem ser obtidos.

Palavras-chave— Reidentificação, Deep Learning, Transfer Learning, Aumentação de Imagens

1 Introdução

As imagens obtidas por câmeras de videomonitoramento têm sido utilizadas em tarefas de vigilância, segurança pública, cuidado de pessoas, etc (Doretto et al., 2011). Em hospitais, creches e asilos permitem o acompanhamento em tempo real de enfermos, crianças e idosos.

Contudo, é na área de segurança pública que esses sistemas têm encontrado sua maior aplicação. A crescente preocupação com ações terroristas nos grandes centros, e o aumento da violência urbana têm elevado os níveis de investimento em sistemas de vigilância por câmeras.

Em sistemas não-automatizados o operador terá que monitorar várias câmeras simultaneamente. Outro desafio é encontrar informações específicas em buscas *off-line*, em meio ao grande volume de imagens capturadas e armazenadas no banco de dados do sistema. Considerando as limitações do ser humano para manter, continuamente, um nível alto de atenção, o monitoramento das imagens pode se tornar um problema quando feito manualmente (Bouma et al., 2013).

A área de visão computacional denominada Reidentificação de Pessoas (*Re-ID*), consiste em verificar se um indivíduo, cuja imagem tenha sido capturada por uma das câmeras de um sistema

de videomonitoramento em um instante de tempo $t1$, corresponde ao indivíduo cuja imagem foi capturada por outra câmera em um instante $t2$, sem que haja sobreposição dos campos de visão das câmeras (Hamdoun et al., 2008).

Quando o problema envolve apenas duas câmeras tem-se a versão *Single Shot* da reidentificação. O objetivo é identificar indivíduos do conjunto de imagens da câmera $c(1)$ entre os indivíduos do conjunto de imagens da câmera $c(2)$. A Figura 1 mostra alguns exemplos de imagens disponíveis na versão *single shot* do dataset PRID2011 (Hirzer et al., 2011).

A abordagem *Multi-Shot*, utiliza imagens de n câmeras ($n > 2$). Nesse caso, o problema pode ser definido como a identificação de um indivíduo cuja imagem foi capturada pelas câmeras $c(1)$ à $c(m)$, ($m < n$), entre os indivíduos das imagens das câmeras $c(m+1)$ à $c(n)$.

De forma geral, em reidentificação as imagens são de baixa qualidade, possuem oclusões, o mesmo indivíduo em diferentes câmeras está em diferentes poses, as câmeras podem apresentar diferenças na calibração e os locais em que estão instaladas podem estar sob diferentes condições de iluminação ambiente (Hamdoun et al., 2008; Hirzer et al., 2011; McLaughlin et al., 2017).

Nesse contexto, esse artigo propõe a utilização



Figura 1: Amostra das imagens do *dataset* PRID2011. As imagens de (1a) à (4a) correspondem à camera A, e as imagens de (1b) à (4b) são as correspondentes imagens da camera B dos mesmos indivíduos.

de técnicas de *Deep Learning* para a reidentificação dos indivíduos do *dataset* PRID2011. Estudos são realizados para verificar a influência dos seguintes fatores no desempenho da reidentificação: (i) aplicação de técnicas de pré-processamento às imagens antes da etapa de reidentificação. (ii) Variação do número de épocas no treinamento da rede neural. (iii) Influência do número de classes e (iv) utilização da aumentação de imagens.

Esse trabalho está constituído pelas seguintes seções: a Seção 2, faz uma revisão bibliográfica do tema de reidentificação de pessoas. A Seção 3 apresenta uma descrição de Deep Learning e Transfer Learning. A Seção 4 faz uma descrição da metodologia utilizada para a execução da pesquisa. A Seção 5 apresenta os resultados obtidos. A Seção 6 relata as principais conclusões. Por fim, a Seção 7 relaciona as referências bibliográficas que embasaram a elaboração do trabalho.

2 Trabalhos Correlatos

Da forma como se apresenta, a reidentificação de pessoas pode ser abordada como um problema de reconhecimento de padrões (Duda et al., 2001), cujo desempenho pode ser associado a estrutura dos descritores utilizados para caracterização dos objetos, e aos algoritmos empregados para estabelecer a correspondência entre objetos desconhecidos e as classes.

No problema de reidentificação cada indivíduo representa uma classe. Dessa forma, a Figura 1 mostra quatro classes de objetos (quatro indivíduos). Uma instância do problema, nesse caso, poderia ser reconhecer o indivíduo 3a da câmera A entre os quatro indivíduos da câmera B.

Um dos problemas relevantes em reidentificação é a representação de pessoas usando descritores robustos, mesmo sob as condições desfavoráveis em que as imagens são obtidas.

2.1 Representação de Pessoas em Reidentificação

Diversas abordagens têm sido propostas para a representação de pessoas em reidentificação. Uma técnica de reidentificação utilizando sequências curtas de vídeo é proposta em (Hamdoun et al., 2008). No esquema proposto, a variabilidade na aparência é capturada com base na análise de pontos de interesse que são extraídos de imagens espalhadas no tempo. Os pontos de interesse identificados em todas as imagens são acumulados durante o período de observação constituindo o descritor que será utilizado na etapa de identificação.

Em (Doretto et al., 2011) é feita uma exposição sobre métodos de reidentificação baseados em descritores de aparência. De acordo com o autor, métodos baseados em aparência podem utilizar informações de apenas uma cena (*single shot*) ou de várias cenas (*multiple shot*). No segundo caso, a fusão de descritores extraídos das diversas cenas pode resultar em melhor desempenho. Outra possibilidade é a associação de descritores locais para se obter uma representação de corpo inteiro, ou uma representação segmentada de regiões do corpo, por exemplo, tronco e membros inferiores.

Entre os métodos baseados em representações do corpo inteiro, pode ser mencionada a *Bag-of-Features*, (Schiele and Crowley, 2000), (Lowe, 2004), (Mikolajczyk and Schmid, 2005), que consiste na representação da imagem por uma coleção de descritores locais, predefinidos de acordo com um dicionário de aparências. Por sua vez, representações de pessoas considerando apenas regiões do corpo extraem descritores baseados em Pontos de Interesse. Métodos baseados em pontos de interesse, podem melhorar a robustez do sistema. Como exemplo de métodos baseados em pontos de interesse pode-se mencionar o *Sift* (Lowe, 2004) e o *Surf* (Bay et al., 2008).

Em (Bouma et al., 2013), são apresentados dois algoritmos de reidentificação denominados *Multi Color-Height Histogram (MCHH)* e *Transform-Normalized Color-Height Histogram*. Baseados no histogramas das imagens, que retêm informações sobre as cores bem como informações espaciais, os métodos apresentam robustez à variações de pose e iluminação. Ambos os métodos são constituídos por duas etapas, o cálculo do histograma, e a verificação da compatibilidade do objeto alvo (pessoa na cena) com um dos objetos rotulados (classes conhecidas).

Em (Khedher et al., 2013) é proposta uma abordagem baseada na representação esparsa de pontos de interesse. Uma estratégia baseada na classificação espectral da silhueta em sequências de imagens é apresentada em (Truong Cong et al., 2010).

Em (Satta, 2013), os autores revisam técnicas utilizadas para a extração de descritores de apa-

rência. Nesse trabalho, o autor classifica os métodos considerando a forma com que a estrutura do corpo é representada e o tipo de característica utilizada para representar a pessoa.

De acordo com (Liu et al., 2014) descritores baseados em cores podem apresentar bom desempenho quando as condições de iluminação entre as cenas não sofrem alterações severas, uma vez que a cor é mais robusta à variações de poses quando comparada à outros tipos de descritores. Descritores estruturais e descritores baseados em textura são mais estáveis sob variações severas de iluminação, contudo, são sensíveis à variações de pose e oclusões. Dessa forma, (Bazzani et al., 2012) e (Perina et al., 2010) propõem a combinação dos dois tipos de descritores para a obtenção de sistemas de reidentificação robustos.

Um método que captura as propriedades estatísticas utilizando a matriz de covariância, e informações de cores no espaço HSV foi proposto por (Zhao et al., 2015). Os descritores, que consistem da combinação do logaritmo da matriz de covariância com o histograma HSV, apresentam robustez a baixa resolução e variações de pose.

2.2 Reidentificação e Deep Learning

Os métodos de reidentificação propostos mais recentemente têm utilizado redes neurais treinadas com algoritmos de deep learning. A capacidade das redes neurais profundas para a extração de descritores e representação das imagens pode melhorar significativamente o desempenho dos sistemas, mesmo considerando os problemas comuns as imagens obtidas por câmeras de videomonitoramento.

Em (McLaughlin et al., 2017) foi proposto um método de reidentificação que utiliza redes neurais profundas nas etapas de extração de descritores e na fase de reconhecimento. Argumentam os autores que a utilização de redes profundas para representar a aparência de pessoas em vídeos, pode extrair descritores mais resistentes aos problemas mais comuns em tarefas de reidentificação. Na etapa de reconhecimento uma segunda rede é treinada para reconhecer pares de imagens similares e não-similares ao objeto alvo.

Em (Xu et al., 2017) os autores enfrentam o problema da pequena quantidade de imagens necessárias para treinar redes profundas, problema que pode levar ao *overfitting*, fazendo uso da técnica de aumento de imagens. A etapa de identificação foi executado utilizando uma arquitetura de *Support Vector Machines*.

3 Fundamentação Teórica

3.1 Redes Neurais Convolutivas

A aprendizagem profunda (do inglês, *Deep Learning*) é um tipo de aprendizado de máquina

no qual um modelo aprende a executar tarefas de classificação diretamente dos dados (*end-to-end task*), sendo geralmente implementada usando uma arquitetura de rede neural. O termo profundo refere-se ao número de camadas na rede, ou seja, quanto mais camadas, mais profunda a rede. As redes neurais tradicionais contêm apenas 2 ou 3 camadas, enquanto as redes profundas podem ter centenas (Bengio, 2009). Neste contexto, destacam-se as Redes Neurais Convolucionais (do inglês, *Convolutional Neural Network - CNN*). Uma Rede Neural Convolucional é uma variação das redes de *perceptron* de múltiplas camadas, tendo sido inspirada no processo biológico de processamentos de dados visuais. De maneira semelhante aos processos tradicionais de visão computacional, uma CNN é capaz de aplicar filtros em dados visuais, mantendo a relação de vizinhança entre os pixels da imagem ao longo do processamento da rede. A Figura 2 ilustra uma CNN tradicional. Este tipo de rede vem sendo amplamente utilizada, principalmente nas aplicações de classificação (Krizhevsky et al., 2012), detecção (Girshick et al., 2013) e reconhecimento em imagens e vídeos (Karpathy et al., 2014).

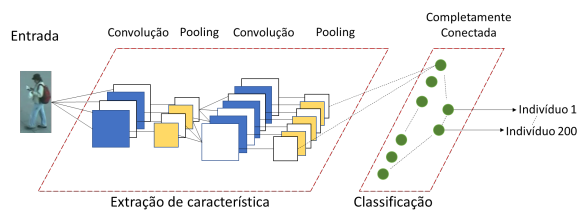


Figura 2: Exemplo de uma Rede Neural Convolucional e suas camadas.

Como definido em (Liu et al., 2017) e analisando-se a Figura 2, uma rede neural convolucional consiste de várias partes com funções diferentes. Uma camada de convolução é composta por diversos neurônios, cada um responsável por aplicar um filtro em um pedaço específico da imagem. De maneira geral, podemos imaginar cada neurônio sendo conectado a um conjunto de *pixels* da camada anterior e que a cada uma dessas conexões se atribui um peso. A combinação das entradas de um neurônio, utilizando os pesos respectivos de cada uma de suas conexões, produz uma saída passada para a camada seguinte. Os pesos atribuídos às conexões de um neurônio podem ser interpretados como uma matriz que representa o filtro de uma convolução de dados no domínio espacial (Liu et al., 2017). Por fim, para a classificação, após o conjunto das camadas de convolução e *pooling* é acrescentada uma camada totalmente conectada, formada normalmente por uma rede *softmax*. Esta camada será responsável por determinar um caminho de decisão a partir das respostas dos filtros vindos das camadas anteriores, para cada classe de resposta.

Assim, diferentemente das técnicas tradicionais de Visão Computacional, nas CNNs é preciso definir somente a arquitetura das camadas: quantidade, tamanho, *stride* e *pooling*. No entanto, em nenhum momento é indicado qual filtro deve ser utilizado, sendo testados diferentes tipos (filtros de escala, angulo, borda, cor, entre outros). Especificamente, o processo de aprendizado da rede altera os pesos ao longo do treinamento até encontrar os melhores valores dos filtros para o conjunto de dados utilizados. Devido ao resultados obtidos, as Redes Neurais Convolucionais têm se tornado o estado-da-arte em diferentes problemas de reconhecimento de padrões, superando em alguns casos a performance humana (Ciresan et al., 2012; Ciresan et al., 2012; Russakovsky et al., 2015; Wu et al., 2016; Kannan et al., 2016).

3.2 Transferência de Aprendizado

Os algoritmos tradicionais de mineração de dados e aprendizado de máquina realizam previsões sobre os dados futuros utilizando modelos estatísticos que são treinados utilizando-se dados de treinamento previamente rotulados (método supervisionado) ou não rotulados (método não supervisionado) (Pan and Yang, n.d.). Já a Transferência de Aprendizado (do inglês, *Transfer Learning*), permite que os domínios, tarefas e distribuições usados no treinamento e no teste sejam diferentes.

No mundo real, observamos muitos exemplos de transferência de aprendizado. Por exemplo, podemos descobrir que aprender a reconhecer maçãs pode ajudar a reconhecer pêras. Da mesma forma, aprender a tocar o órgão eletrônico pode facilitar o aprendizado do piano. O estudo da transferência de aprendizagem é motivado pelo fato de que as pessoas podem aplicar, de forma inteligente, o conhecimento aprendido anteriormente para resolver novos problemas mais rapidamente ou com melhores soluções (Pan and Yang, n.d.). De outra maneira, a transferência de aprendizado é uma abordagem que aplica o conhecimento de um tipo de problema a um problema diferente, mas relacionado. Neste caso, simplesmente troca-se a última camada da rede (*fully-connected layer*) e realiza-se o treinamento novamente, porém, com novos dados, como mostra a Figura 3.

Como já definido por (Pan and Yang, n.d.), na transferência de aprendizagem, temos as três principais questões de pesquisa a seguir: (1) O que transferir; (2) Como transferir; e (3) Quando transferir.

“O que transferir” pergunta qual parte do conhecimento pode ser transferida entre domínios ou tarefas. Algum conhecimento é específico para domínios ou tarefas individuais, e algum conhecimento pode ser comum entre domínios diferentes, de modo que eles podem ajudar a melhorar

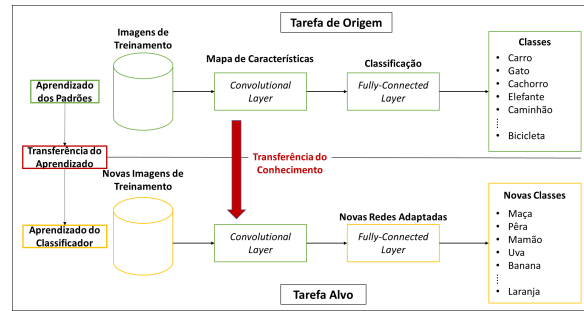


Figura 3: Realização da Transferência de aprendizado.

o desempenho do domínio ou tarefa de destino. Depois de descobrir qual conhecimento pode ser transferido, algoritmos de aprendizado precisam ser desenvolvidos para transferir o conhecimento, o que corresponde à questão de “como transferir”.

“Quando transferir” pergunta em quais situações a transferência de aprendizado deve ser feita. Da mesma forma, é interessante saber em que situações o conhecimento não deve ser transferido. Assim, em algumas situações, quando os domínios de origem e de destino não estão relacionados entre si, a transferência de aprendizado pode não ter êxito. Na pior das hipóteses, pode até prejudicar o desempenho da aprendizagem no domínio-alvo, uma situação que é muitas vezes referida como transferência negativa.

É comum que trabalhos sobre transferência de aprendizado se concentram em “O que transferir” e “Como transferir”, assumindo implicitamente que os domínios de origem e de destino estejam relacionados entre si. No entanto, tal suposição muitas vezes não é verdade em problemas reais. Neste sentido, este trabalho busca avaliar a utilização da transferência de aprendizado, comparando sua performance com uma CNN tradicional para um problema de re-identificação, como descrito nas próximas seções.

4 Metodologia proposta

Para os experimentos foi utilizado o *dataset* PRID2011 (Hirzer et al., 2011), que suporta as versões *single-shot* e *multi-shot* do problema de reidentificação. O conjunto para a versão *single-shot* é constituída por 1.134 imagens, sendo 385 para câmera A e 749 para a câmera B. O *multi-shot* é composto por 96.121 imagens, sendo 43.862 para a câmera A e 52.259 para a câmera B. Neste trabalho foram utilizadas as imagens correspondentes às primeiras 200 classes presentes em ambos subconjuntos.

Na etapa de reidentificação, em ambos os casos (*single-shot* e *multi-shot*) foram utilizadas redes neurais convolucionais *CNN* e a rede AlexNet (Krizhevsky et al., 2012), essa última foi reutilizada com o uso da técnica de *transfer learning*.

O treinamento de redes neurais profundas requer uma grande quantidade de imagens. Assim, na reidentificação versão *single-shot*, foi utilizada a técnica de aumento para obter um número de imagens que tornassem os experimentos viáveis. As imagens aumentadas foram geradas a partir das originais, sobre as quais foram aplicadas alguns tipos de transformações.

A etapa experimental consistiu na realização das seguintes investigações:

Inicialmente foi verificado o efeito da aplicação de técnicas de pré-processamento às imagens antes da reidentificação. As seguintes técnicas foram investigadas: equalização de histogramas, ajuste no mapa de cores, transferência de cores para uma imagem de referência e correção automática de branco.

Em seguida foi avaliado se a reutilização de redes neurais através do uso da técnica de *transfer learning* seria uma opção vantajosa quando comparada ao projeto completo de uma nova rede neural convolucional. O objetivo desse teste é verificar o número de épocas necessárias para o treinamento da nova rede, e o seu efeito sobre o desempenho da reidentificação.

No terceiro experimento foi estudado o impacto da quantidade de classes envolvidas no problema. Para esse teste o desempenho da reidentificação foi avaliado para as seguintes quantidades de classes: 10, 50, 100, 150 e 200 classes.

No quarto e último teste, foi investigado o efeito da aumento de imagens sobre o conjunto *single-shot*. As seguintes operações de transformação foram realizadas sobre as imagens originais: rotação, espelhamento horizontal, ampliação e deslocamento. A partir da combinação aleatória destas técnicas foram geradas 99 imagens adicionais para cada imagem original disponível no conjunto *single-shot*, ampliando as imagens inicialmente consideradas de 400 para 40.000.

A avaliação do desempenho da reidentificação em todos os experimentos foi realizada com a utilização da curva CMC (*Cumulative Matching Characteristic*). A curva CMC indica a taxa de acerto na reidentificação (eixo Y) de acordo com o *rank* considerado (eixo X). Nesse contexto, *Rank* é definido como o número de imagens dentre as quais espera-se obter identificação correta. De acordo com a curva CMC, os melhores desempenhos correspondem à taxas de acerto mais elevadas em *rank*s mais baixos.

Todos os códigos foram implementados em Matlab versão 2017a e ficará disponibilizado à comunidade ao fim das atividades.

5 Resultados

Nessa seção são apresentados os resultados dos quatro experimentos descritos na metodologia sobre os conjuntos *Multi-Shot* e *Single-Shot*.

5.1 Abordagem Multi-Shot

Considerando que a versão *Multi-Shot* do conjunto de dados é mais adequado à aplicação de técnicas de *deep learning* pelo maior número de imagens disponibilizadas, esse conjunto foi escolhido para os testes iniciais.

5.1.1 Pré-processamento

Foram realizados diversos testes utilizando as técnicas de pré-processamento propostas na metodologia. Dentre as técnicas de pré-processamento aplicadas, a correção de branco foi a que apresentou melhor resultado, tendo sido, portanto, a utilizada na sequência do trabalho.

A Figura 4 mostra a comparação dos resultados obtidos com as imagens, com e sem correção de branco, obtidos por uma rede neural treinada pelos pesquisadores e outra reutilizada com o uso do *transfer learning*.

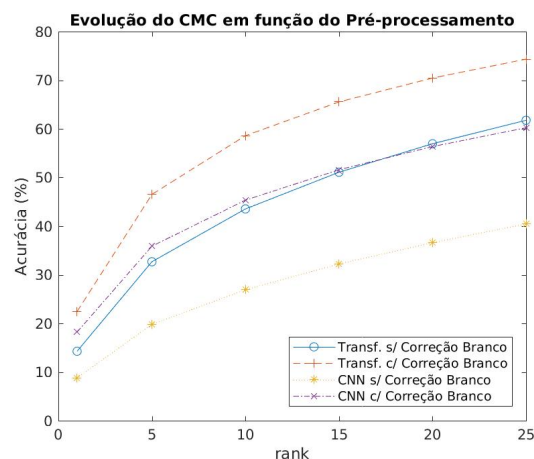


Figura 4: Efeito da aplicação de pré-processamento, antes da entrada das imagens para o treinamento.

Os resultados mostram que a correção de branco, com ambas as redes, melhorou o resultado da reidentificação. Pode-se notar também que a reutilização de redes já treinadas é uma opção viável, uma vez que elimina o tempo necessário ao treinamento dessas topologias de rede, que nos problemas de reidentificação podem ser de algumas horas, ou mesmo dias.

De forma geral, a utilização de técnicas de pré-processamento ou a combinação de várias delas podem melhorar os resultados da reidentificação.

5.1.2 Épocas

A Figura 5 mostra os resultados do experimento para verificação do efeito do número de épocas na fase de treino sobre o desempenho do sistema. Nota-se que a melhora no desempenho aumenta com o número de épocas de treinamento da rede.

Contudo, o treinamento de redes neurais profundas é computacionalmente custoso, o que pode requerer poderosos hardwares e longos períodos de treinamento para um número elevado de épocas.

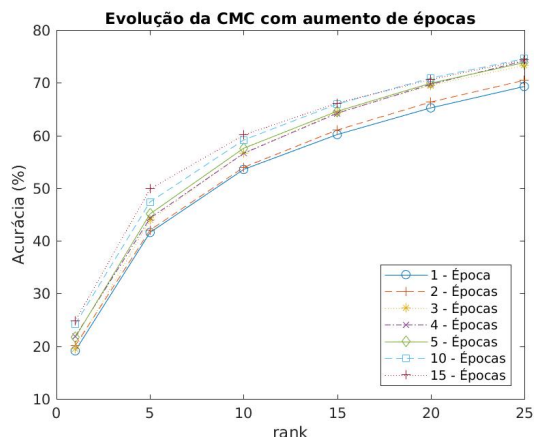


Figura 5: Efeito da variação da quantidade de épocas durante o treinamento.

5.1.3 Classes

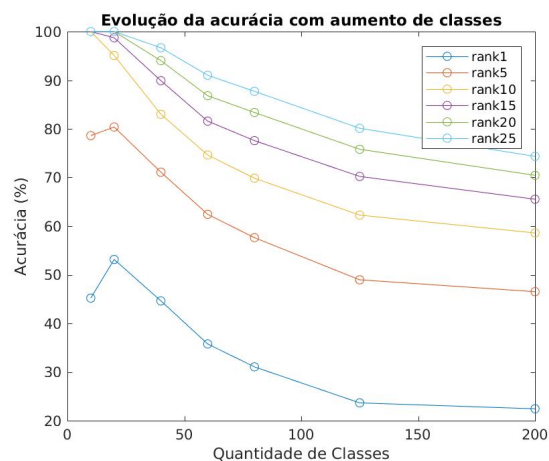
O resultado dos experimentos para verificar a influência do número de classes estão mostrados na Figura 6. Na Figura 6a observa-se que a taxa de acerto na reidentificação para o conjunto PRID2011 decresce abruptamente a partir de um número de classes (em torno de 20), para todos os ranks. A Figura 6b mostra que para todos os ranks, as maiores taxas de acerto são obtidas para menores quantidades de classes envolvidas no problema. Esse resultado pode auxiliar na determinação do número de classes quando da configuração de algoritmos de *ensemble*.

As Tabelas 1 e 2 apresenta os resultados de análises do *F1 score* das cinco piores e das cinco melhores classes em termos da taxa de acerto.

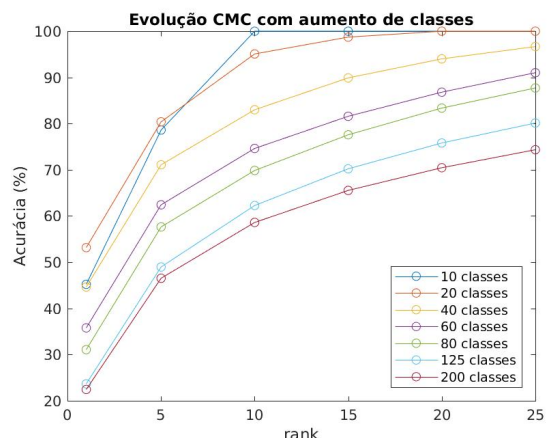
Tabela 1: Cinco classes com os piores valores de *F1 score*

Classe	<i>F1 score</i>
9	0
10	0
19	0
21	0
25	0

Dentre as classes com piores estatísticas, o valor 0 (zero) indica que nenhum dos exemplares da classe no conjunto foi reconhecido corretamente, considerando *rank* 1. Na condição considerada, o conjunto utilizado contém 32 classes. Esse resultado pode indicar a necessidade de um trabalho mais aprofundado em relação às imagens dessas classes visando a definir técnicas de pré-processamento mais adequadas a fim de diminuir



(a) degradação dos ranks



(b) degradação das curvas CMC

Figura 6: Efeito da quantidade de classes no resultado do classificador.

Tabela 2: Cinco classes com os melhores valores de *F1 score*

Classe	<i>F1 score</i>
162	0,8750
191	0,8815
107	0,9240
110	0,9360
16	0,9699

o impacto negativo destas classes sobre o resultado geral da identificação.

A análise das matrizes de confusão obtidas com a utilização dos dois tipos de rede (rede reutilizada e rede treinada pelos autores), mostra que as classes com piores estatísticas, nos dois casos, são diferentes. Notamos ainda uma diferença das classes com piores estatísticas, conforme a técnica utilizada. Este resultado indica a possibilidade de melhoria dos resultados via combinação de classificadores.

5.2 Single-Shot

A reidentificação no caso *single shot* apresentou-se ainda mais desafiadora, pois além do problema relacionado à qualidade, já discutido, temos que lidar ainda com a pequena quantidade de imagens para cada indivíduo, o que não favorece a utilização de técnicas de *deep learning*.

5.2.1 Aumentação de imagens

Para esses testes, a partir de cada exemplar original do conjunto PRID2011 *Single-Shot*, pela aplicação das transformações indicadas previamente, foram obtidos conjuntos com 5, 10, 25, 50, 75 e 100 imagens. A Figura 7 que ilustra os resultados da aumentação, mostra que o desempenho da reidentificação é melhorado, até um certo limite, a medida que a quantidade de imagens por classe é aumentada. Contudo deve-se atentar para que a aumentação não seja utilizada de forma inadequada, pois o aumento do número de imagens por classe, eleva também o tempo de processamento.

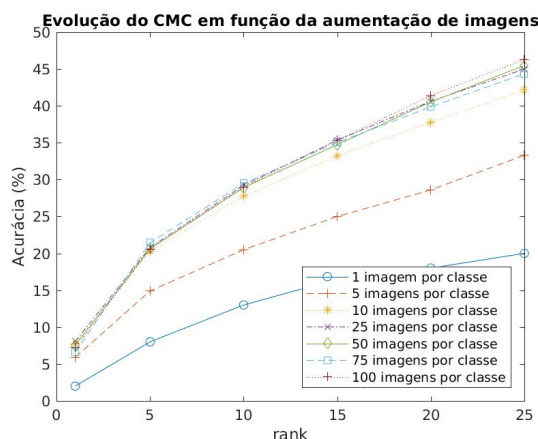


Figura 7: Efeito da aumentação de imagens para a reidentificação a partir de um *dataset single shot*.

6 Conclusões e Trabalhos Futuros

Neste trabalho foram estudadas técnicas de *Deep Learning* aplicadas à reidentificação de pessoas. Foi analisada a influência sobre o desempenho final do pré-processamento das imagens, do número de épocas de treinamento da rede, do número de classes utilizadas e o uso de aumentação de imagens.

Embora, a quantidade de imagens no conjunto Prid2011 *Multi-Shot* (40.033) seja suficiente, as mesmas se apresentam com um alto grau de desbalanceamento. Por exemplo, existem classes com 5 imagens, enquanto outras apresentam 179, o que pode comprometer o desempenho do sistema. Em relação ao pré-processamento, verificou-se que sua utilização pode melhorar o desempenho da reidentificação.

A exploração dos parâmetros da rede permitiu identificar pontos de atenção para uma melhora dos resultados no futuro. Foi verificado ainda a aumentação de imagens para o conjunto *single-shot*, o que trouxe uma melhora de performance para os classificadores gerados. Revelou-se também como um caminho viável, podendo ser utilizado no tratamento do desbalanceamento entre as classes. Pode-se expandir as transformações aplicadas, e o encadeamento de outras técnicas, para simular os problemas de calibração de cores das câmeras, e iluminação, trazendo um ganho adicional para a reidentificação.

Foi possível isolar as classes que contribuem negativamente para o desempenho dos classificadores, abrindo novas linhas de investigação para serem exploradas. Com um tratamento mais direcionado, espera-se diminuir o impacto destas classes.

Trabalhos futuros irão explorar todas estas hipóteses, buscando tratamento específico para os pontos mais problemáticos. Será projetado também um classificador a partir da técnica de *ensemble*, confrontando com os resultados já obtidos, bem como os disponibilizados pela comunidade acadêmica.

Referências

- Bay, H., Ess, A., Tuytelaars, T. and Gool, L. V. (2008). Speeded-up robust features (surf), *Computer Vision and Image Understanding* **110**: 346–359.
- Bazzani, L., Cristani, M., Perina, A. and Murino, V. (2012). Multiple-shot person re-identification by chromatic and epitomic analyses, *Pattern Recogn. Lett.* **33**(7): 898–903.
- Bengio, Y. (2009). Learning Deep Architectures for AI, *Foundations and Trends® in Machine Learning* **2**(1): 1–127.
- Bouma, H., Borsboom, S., den Hollander, R. J. M., Landsmeer, S. H. and Worring, M. (2013). Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination, *Proceedings [SPIE SPIE Defense, Security, and Sensing]*, SPIE.
- Ciresan, D., Meier, U. and Schmidhuber, J. (2012). Multi-column deep neural networks for image classification, *IEEE*, pp. 3642–3649.
- Doretto, G., Sebastian, T., Tu, P. H. and Rittscher, J. (2011). Appearance-based person re-identification in camera networks: problem overview and current approaches, *J. Ambient Intelligence and Humanized Computing* **2**: 127–151.

- Duda, R. O., Hart, P. E. and Stork, D. G. (2001). *Pattern Classification (2nd Ed)*, Wiley.
- Girshick, R. B., Donahue, J., Darrell, T. and Malik, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation, *CoRR abs/1311.2524*.
- Hamdoun, O., Moutarde, F., Stanculescu, B. and Steux, B. (2008). Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences, *Proceedings 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC-08)*.
- Hirzer, M., Beleznai, C., Roth, P. M. and Bischof, H. (2011). Person Re-Identification by Descriptive and Discriminative Classification, *Proc. Scandinavian Conference on Image Analysis (SCIA)*.
- Kannan, A., Kurach, K., Ravi, S., Kaufmann, T., Tomkins, A., Miklos, B., Corrado, G., Lukacs, L., Ganea, M., Young, P. and Ramavajjala, V. (2016). Smart reply: Automated response suggestion for email, *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, ACM, New York, NY, USA, pp. 955–964.
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L. (2014). Large-Scale Video Classification with Convolutional Neural Networks, *IEEE*, pp. 1725–1732.
- Khedher, M. I., Yacoubi, M. A. E. and Dorizzi, B. (2013). Multi-shot surf-based person re-identification via sparse representation, *Proceedings 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, in F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger (eds), *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., pp. 1097–1105.
- Liu, C., Gong, S., Loy, C. C. and Lin, X. (2014). Evaluating feature importance for re-identification, *Person Re-Identification*.
- Liu, J., Gao, X., Bao, N., Tang, J. and Wu, G. (2017). Deep convolutional neural networks for pedestrian detection with skip pooling, *IEEE*, pp. 2056–2063.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* **60**: 91–110.
- McLaughlin, N., Rincon, J. M. D. and Miller, P. C. (2017). Person reidentification using deep convnets with multitask learning, *IEEE Transactions on Circuits and Systems for Video Technology* **27**: 525–539.
- Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Learning* **27**: 1615–1630.
- Pan, S. J. and Yang, Q. (n.d.). A survey on transfer learning, **22**(10): 1345–1359.
- Perina, A., Murino, V., Cristani, M., Farenzena, M. and Bazzani, L. (2010). Person re-identification by symmetry-driven accumulation of local features, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR)*, Vol. 00, pp. 2360–2367.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge, *Int. J. Comput. Vision* **115**(3): 211–252.
- Satta, R. (2013). Appearance descriptors for person re-identification: a comprehensive review, *CoRR abs/1307.5748*.
- Schiele, B. and Crowley, J. L. (2000). Recognition without correspondence using multidimensional receptive field histograms, *Int J Comput Vis* **36**: 31–50.
- Truong Cong, D. N., Khoudour, L., Achard, C., Meurie, C. and Lezoray, O. (2010). People re-identification by spectral classification of silhouettes, *Signal Process.* **90**(8): 2362–2374.
- Wu, Y., Schuster, M., Chen, Z., V. Le, Q., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, u., Gouws, S., Kato, Y., Kudo, T., Kazawa, H. and Dean, J. (2016). Google’s neural machine translation system: Bridging the gap between human and machine translation.
- Xu, M., Tang, Z., Yao, Y., Yao, L., Liu, H., Xu, J. N. and Wah, C. (2017). Deep learning for person reidentification using support vector machines, *Advances in Multimedia*.
- Zhao, Y., Wang, L., Zhao, X. and Liu, Y. (2015). Multi-shot person re-identification by features learned from the third-party image set, *KSII Transactions on Internet e Information Systems* **9**: 775–792.