EDITORA CONHECIMENTO LIVRE

SISTEMA EMBARCADO PARA IDENTIFICAÇÃO DE ALUNOS UNIFORMIZADOS UTILIZANDO VISÃO COMPUTACIONAL NO RASPBERRY PI

Gabriela Gomes dos Santos
Luís Alexandre Ferreira Bueno
Ruy de Oliveira
Valtemir Emerencio do Nascimento

Gabriela Gomes dos Santos
Luís Alexandre Ferreira Bueno
Ruy de Oliveira
Valtemir Emerencio do Nascimento

SISTEMA EMBARCADO PARA IDENTIFICAÇÃO DE ALUNOS UNIFORMIZADOS UTILIZANDO
VISÃO COMPUTACIONAL NO RASPBERRY PI

1ª ed.

Piracanjuba-GO
Editora Conhecimento Livre
Piracanjuba-GO

1ª ed.

Dados Internacionais de Catalogação na Publicação (CIP)

Santos, Gabriela Gomes dos
S237S SISTEMA EMBARCADO PARA IDENTIFICAÇÃO DE ALUNOS
UNIFORMIZADOS UTILIZANDO VISÃO COMPUTACIONAL NO RASPBERRY PI
/ Gabriela Gomes dos Santos. Luís Alexandre Ferreira Bueno. Ruy de Oliveira. Valtemir Emerencio
do Nascimento. – Piracanjuba-GO
Editora Conhecimento Livre, 2021
24 f.: il
DOI: 10.37423/2021.edcl379
ISBN: 978-65-89955-98-6
Modo de acesso: World Wide Web
Incluir Bibliografia
1. dectecção-de-objetos- 2. opencv 3. raspberrypi 4. visão-computacional I. Santos, Gabriela
Gomes dos II. Bueno, Luís Alexandre Ferreira III. Oliveira, Ruy de IV. Nascimento, Valtemir
Emerencio do V. Título

CDU: 4

<https://doi.org/10.37423/2021.edcl379>

O conteúdo dos artigos e sua correção ortográfica são de responsabilidade exclusiva dos seus respectivos autores.

EDITORA CONHECIMENTO LIVRE

Corpo Editorial

Dr. João Luís Ribeiro Ulhôa

Dra. Eyde Cristianne Saraiva-Bonatto

Dr. Anderson Reis de Sousa

MSc. Frederico Celestino Barbosa

MSc. Carlos Eduardo de Oliveira Gontijo

MSc. Plínio Ferreira Pires

Editora Conhecimento Livre

Piracanjuba-GO

2021

10.37423/2021.edcl379

Resumo: A segurança de instituições e empresas é um assunto cada vez mais discutido, e tem levado à criação de políticas de acesso cada vez mais rigorosas. Um dos métodos de controle de acesso mais comuns e eficazes é o uso de uniformes. Contudo, a verificação do porte de fardas manualmente implica na necessidade de recursos humanos e é sujeita a falhas. Mecanismos de Inteligências Artificiais baseadas em Redes Neurais são reconhecidamente eficientes para o reconhecimento de padrões em várias áreas, e quando aliadas a ferramentas de processamento de imagens, tornam-se bem apropriadas para a identificação de objetos em vídeos e/ou fotografias com alta precisão. Este trabalho propõe um sistema automatizado, portátil e de baixo custo para identificar alunos uniformizados e permitir-lhes acesso ao ambiente escolar. O sistema captura as imagens dos alunos e as avalia com base em conceitos modernos de Visão Computacional. A solução proposta faz uso de *software open source* para implementar algoritmos de uma Rede Neural Convolucional embarcados num computador de baixo custo *Raspberry Pi*. Os resultados obtidos nas avaliações preliminares são animadores.

Palavras-Chave: Redes Neurais Convolucionais, Visão Computacional, Sistema de Monitoramento.

1. INTRODUÇÃO

Estudos realizados pelo Instituto de Ciências da Educação dos Estados Unidos – IES [1], apontaram que o uso de uniforme obrigatório está sendo adotado em escolas como medida de monitoramento de acesso: pessoas não fardadas se destacam na multidão, ajudando a segurança do ambiente identificar rapidamente invasores e pessoas que não devem ter acesso ao local.

Para avaliar se os alunos atendem ao padrão de vestimenta da instituição, são realizadas verificações manuais, o que exige recursos humanos e é sujeita a falhas. A falta de funcionários no âmbito escolar do estado de São Paulo levou a instalação de câmeras de vigilância, com a intenção de auxiliar neste processo. No entanto o uso de câmeras de segurança provou-se ineficaz, pois além do alto custo dos sistemas proprietários atuais, é necessário realocar pessoal para fazer o monitoramento das câmeras, e a falta de funcionários continuou sendo um empecilho para o aumento o monitoramento da escola [2].

Um meio de diminuir a necessidade de recursos humanos para monitoramento é através da automatização do processo por meio de técnicas de visão computacional, as quais permitem ao computador identificar e classificar objetos, ações e pessoas em imagens e vídeos, e neste caso, como ferramenta de segurança para instituições públicas e privadas [3]. Um sistema de monitoramento por imagem eficaz deve ser capaz de detectar e reconhecer objetos ou pessoas, e a partir disso verificar se as vestimentas atendem ou não aos padrões da instituição, e caso contrário gerar um alerta para que seja verificada fisicamente a situação do indivíduo [4].

Diversas abordagens têm sido propostas para solucionar este problema, e uma delas é o *Deep Learning*, capaz de processar e analisar grandes quantidades de dados. Nela, são utilizadas Redes Neurais Convolucionais (CNNs) as quais realizam tarefas de reconhecimento visual, como classificação, localização e detecção de objetos em imagens [5][6][7].

Este trabalho propõe utilizar técnicas de processamento de imagem com Redes Neurais Convolucionais para realizar a identificação de alunos uniformizados, a fim de controlar o acesso de pessoas de uma instituição, capaz de ser executado com baixos requisitos de Hardware como o microcomputador *Raspberry Pi 3*. A flexibilidade e o baixo custo do *Raspberry Pi*, juntamente com o desempenho e a precisão da CNN possibilitam o desenvolvimento de um sistema automatizado para monitoramento de alunos uniformizados com baixos investimentos.

Desta forma, as demais seções deste artigo estão estruturadas como segue: na seção 2 é apresentado o estado da arte da área de utilização de Visão Computacional para segurança de ambientes privados. A seção 3 descreve os principais conceitos da construção de um sistema de localização e detecção de objetos. Na seção 4 é mostrado o método proposto por este trabalho, na seção 5 estão os resultados obtidos, e, por fim a seção 6 apresenta as conclusões deste trabalho e trabalhos futuros.

2. TRABALHOS RELACIONADOS

Nos últimos anos, diversos pesquisadores têm mostrado interesse em mecanismos de Inteligência Artificial visando solucionar problemas de autenticação nos sistemas de monitoramento de instituições públicas e privadas. Visão Computacional tem sido amplamente utilizada em sistemas de monitoramento, como mecanismo de autenticação [8], [9], [10], [11], [12] e de detecção de presença de pessoas [13], [14], [15].

Os autores em [9] expõem um protótipo de reconhecimento facial baseado em um estudo feito sobre o algoritmo de Viola Jones, que foi desenvolvido especificamente para o treinamento de redes neurais para reconhecimento de rostos de pessoas. Sua principal característica é a detecção de características através de funções de *Haar*. Esse trabalho aplica o algoritmo Viola Jones em uma rede neural do tipo *MultiLayer Perceptron* (MLP) para testá-lo em condições não favoráveis, como má iluminação e câmera de baixa resolução, utilizando um dispositivo *Raspberry Pi 3*. Apesar dos autores definirem o algoritmo Viola Jones como “sendo um dos algoritmos mais eficientes em termos de desempenho e processamento, por não requerer grande poder computacional”, não foi possível realizar o devido treinamento, em razão do baixo poder computacional do equipamento. Nestas condições, houve uma taxa de acerto de aproximadamente 50%, mas não há embasamento matemático suficiente para replicação do mesmo.

Os autores em [16] propõem um método de identificação de roupas, com base na geometria computacional *Constrained Delaunay Triangulation* (CDT), sem modelo de vestuário pré-definido, guiado apenas pelo rosto da pessoa. Para realizar a detecção de peças de roupa, os autores utilizaram o detector facial já implementado no *framework OpenCV*, localizando o torso da pessoa. Verificou-se que diferentes iluminações atrapalham o reconhecimento das cores, atrapalhando o algoritmo de corte de grafos, e por isso utilizaram a própria CDT para suprir os ruídos nas imagens, criando um método de supressão. Por fim, o método mostrou-se eficaz na detecção das roupas, dependendo da

detecção facial para o reconhecimento do tronco da pessoa. Se o rosto da pessoa não for detectado e se a pessoa detectada possui manchas na pele, o método falha.

Yang e Yu [17], desenvolveram um sistema de reconhecimento de roupas em tempo real em vídeo de vigilância. O sistema reconheceu 8 tipos diferentes de vestimentas, utilizando o método de segmentação de roupas, avaliando vários recursos de textura baseado no *Histogram of Oriented Gradients* (HOG), *Bag-of-Words* (BOW) *Features*, e respostas DCT. Cada categoria de roupas é controlada por um vetor SVM (Máquina de Vetores de Suporte) linear. O sistema utiliza um histograma de cores para identificar as áreas do corpo, realizando proporções dos membros, estimativas de idade e sexo. O sistema atinge taxa de recuperação média de 80% em FPR 0,1.

Por fim é possível notar uma vasta quantidade de trabalhos cujo foco é o monitoramento de ambientes através da detecção de pessoas em vídeo, seja por movimento ou pela vestimenta do indivíduo. Redes Neurais Artificiais mostraram-se capazes de monitorar o ambiente sem a necessidade de recursos humanos. Tais resultados motivaram o desenvolvimento deste trabalho como proposta de reconhecimento de alunos uniformizados em uma instituição de ensino, baseado nos avanços recentes da visão computacional.

3. SISTEMAS DE LOCALIZAÇÃO E CLASSIFICAÇÃO DE OBJETOS

Integrando as áreas de processamento digital de imagens e inteligência artificial, a visão computacional tem como objetivo desenvolver algoritmos capazes de identificar, diferenciar e classificar imagens [3]. O Aprendizado de Máquina – do inglês *Machine Learning* – permite que os algoritmos, a partir de uma base de dados, aprendam como executar uma tarefa, como por exemplo, o reconhecimento de objetos em imagens [18].

Esta seção descreve os principais conceitos utilizados para desenvolver o sistema para realizar a identificação de alunos uniformizados.

3.1 REDES NEURAI

Segundo Haykin [19], Sistemas de Inteligência Artificial são ferramentas capazes de realizar tarefas sem que haja uma definição explícita prévia do algoritmo a ser seguido. Neurônios são as unidades de processamento fundamentais para a operação de uma Rede Neural. Cada neurônio possui um conjunto de sinais de entrada combinados a uma matriz de pesos seguidos por uma constante denominada bias, conforme apresentado na Figura 1(a).

Como demonstrado na Equação 1, no neurônio é aplicado uma operação matemática capaz de mapear o conjunto de informações de entrada ao conjunto de saída por meio de uma função de ativação [20].

$$y = f \left[\left(\sum_i E_i * W_i \right) + b \right] \quad (1)$$

Conforme ilustrado na Figura 1(b), os Neurônios de uma Rede Neural Artificial são organizados em grupos com funções específicas, denominados camadas, cujas entradas são informações provenientes da camada anterior [21]. A primeira camada é denominada Camada de Entrada, e esta recebe como entrada dados externos para serem processados. As camadas subsequentes são chamadas Camadas Ocultas, e não há como determinar a função específica de cada uma delas. Nelas são realizadas a extração de características dos dados de entrada. A última camada, chamada Camada de Saída, recebe as informações das Camadas Ocultas e emite uma resposta.[22].

Figura 1 - Arquitetura de Redes Neurais Artificiais: (a) estrutura dos neurônios e (b) estrutura das camadas

3.2 REDES NEURAIS CONVOLUCIONAIS

Redes Neurais Convolucionais (CNNs) possuem estrutura semelhante à das Redes Neurais Artificiais (ANNs), sendo constituídas de múltiplas camadas interligadas. A principal diferença entre CNNs e ANNs está na organização de suas camadas, onde CNNs não possuem somente camadas totalmente conectas, e sim se dividem em camadas de convoluções e de subamostragem. CNNs se mostraram mais eficientes que ANNs para reconhecimento de padrões e segmentação de imagens, detecção e classificação de objetos, pois requerem menor poder computacional que as ANNs [23].

Uma das técnicas utilizadas para a detecção e classificação de objetos em uma imagem é chamada *Deep Learning*, e se baseia em Redes Neurais Convolucionais, as quais possuem uma arquitetura multicamadas [24], como mostrada na Figura 2.

Figura 2 - Arquitetura de Redes Neurais Convolucionais

Uma Rede Neural Convolucional pode possuir camadas de convolução, camadas de subamostragem, camadas de normalização de contraste e camadas completamente conectadas. Uma CNN é composta por uma ou mais sequências destas camadas. A presença de camadas convolucionais não totalmente conectadas é a principal característica de uma CNN, que tem como função a extração de características dos dados recebidos na entrada. A saída produzida por essa operação são filtros de convolução (matrizes), chamados também de mapas de características da imagem, que armazenam os pesos das conexões entre os neurônios para realizar a detecção e classificação do objeto [21].

Outra característica presente em CNNs é a camada de subamostragem (*pooling layer*), que tem por objetivo diminuir a dimensionalidade do mapa de características agregando os pesos contidos na camada anterior produzindo, assim, outro mapa de características, porém resumido. As principais formas desta subamostragem são: selecionar o *max pooling* (como mostra a Figura 3a) mantendo o maior peso de característica e *average pooling* (como mostra a Figura 3b) fazendo a média dos pesos [25].

Figura 3 - Exemplos de Subamostragem com filtro 2x2: (a) *max pooling* e (b) *average pooling*

A camada de normalização de contraste local (LCN) é uma camada opcional nas Redes Neurais Convolucionais. Quando utilizada esta fica na saída da camada de subamostragem, normalizando o contraste da imagem localmente, ou seja, em cada pixel da imagem [26]. Por fim a camada completamente conectada, que significa que todos os neurônios da camada anterior estão conectados com os da camada seguinte, pode ser adicionada mais de uma vez antes da saída em uma CNN, gerando descritores de característica da imagem para a camada de saída [21]. A última camada da CNN, também totalmente conectada, é responsável por realizar a classificação dos dados [24].

3.3 FUNÇÕES DE ATIVAÇÃO

Funções de ativação são parte essencial da arquitetura de uma rede neural, pois adicionam a característica principal desta abordagem: não-linearidades. Sem esta característica, Redes Neurais Convolucionais seriam simples regressões lineares [27].

Assim como camadas convolucionais, camadas de ativação são filtros. Entretanto, o seu núcleo é unidimensional, i.e., é aplicado sobre cada elemento da matriz de características individualmente. Seu propósito é multiplicar cada um dos elementos do mapa de características por uma função específica, chamada função de ativação, com a intenção de indicar se este elemento é relevante ou não para a resposta final [28].

Uma função de ativação simples é o degrau unitário, onde respostas menores que zero não ativam o neurônio e respostas maiores que zero o ativam. O gradiente tem como função indicar a direção de minimização do erro, e o grande problema desta função de ativação é seu gradiente nulo, que impede o cálculo de erros. Sendo assim, funções não lineares deriváveis (ou seja, sem pontos de descontinuidade), são as melhores escolhas para tal tarefa [19].

Até pouco tempo, a função *Sigmóide*, apresentada na Equação 2, era a mais comumente utilizada pois assume somente valores binários e possui uma transição contínua, fazendo sua derivada sempre menor que 1. Derivadas muito pequenas impactam a propagação dos gradientes para escolha dos pesos da rede, e por isto seu uso não é mais recomendado [20].

$$\sigma(x) = \frac{1}{1 + e^x} \quad (2)$$

Uma substituta proposta a função *Sigmóide* é a *Tangente Hiperbólica*, *tanh*, apresentada na Equação 3. Como seu intervalo de valores varia entre -1 e 1, sua derivada é um pouco maior que a da função *Sigmóide*, e deste modo não prejudica o gradiente [19], que é responsável pela atualização dos pesos e bias da rede no sentido de levar a rede neural à convergência.

$$\tanh(x) = 2\sigma(2x) - 1 \quad (3)$$

Diferente das funções *Sigmóide* e *Tangente Hiperbólica*, a função *Unidade de Retificação Linear (ReLU)* é naturalmente não linear, e seu intervalo varia de 0 a infinito. Como a saída desta função é constante para entradas menores do que zero, o gradiente dela também é 0, o que pode prejudicar a ativação dos neurônios. Surge então a função *ReLU6*, apresentada na Equação 4, a mais utilizada atualmente [28], cuja principal característica é adicionar uma pequena inclinação negativa para entradas menores que 0.

$$ReLU6(x) = \min(\max(x, 0), 6) \quad (4)$$

A função *Swish*, apresentada na Equação 5 é graficamente similar a *ReLU* original, com limites inferiores e sem limites superiores. A maioria das funções de ativação é monótona, ou seja, seu valor nunca diminui à medida que a entrada aumenta. A *Swish* possui uma inversão de sua derivada logo antes do zero, tornando-a suave e não monótona. [29].

$$swish(x) = x * sigmoid(x) \quad (5)$$

4. SISTEMA DE DETECÇÃO PROPOSTO

A figura 4 ilustra o funcionamento do sistema proposto por este trabalho. De acordo com o que está mostrado na Figura 4, o mecanismo proposto é extremamente simples e exige apenas uma câmera de

vídeo e um *Raspberry Pi*, como servidor, para prover auxílio no controle da catraca de acesso, mediante reconhecimento de alunos uniformizados.

Figura 4 - Ilustração do sistema de detecção e liberação de acesso de aluno uniformizado

Quando um estudante tenta adentrar a instituição com seu cartão de acesso, é capturada uma imagem na câmera, que será analisada pelo servidor, no *Raspberry Pi*, que executa o algoritmo desenvolvido para verificar o uso do uniforme. O acesso à instituição é liberado apenas após uma resposta positiva do servidor.

Pode-se resumir o sistema para identificação de alunos uniformizados em duas etapas principais: verificar se o aluno está uniformizado e confirmar através da localização do logotipo no uniforme, demonstrado no diagrama de blocos da Figura 5.

Figura 5 - Modelo do Sistema de Identificação de Alunos Uniformizados

A seguir, cada um dos blocos do Método Proposto é apresentado detalhadamente.

4.1 VERIFICAÇÃO DE UNIFORME

A primeira etapa do método proposto consiste em identificar o uniforme. Para isto é utilizada uma Rede Neural Convolucional que localiza e classifica o formato da camisa do aluno. O uniforme pode estar localizado em qualquer lugar da imagem, então faz-se necessário a criação de caixas delimitadoras em regiões possíveis de localização do objeto. Para demarcar as áreas de interesse, utilizou-se uma Rede Neural Convolucional delimitadora [30], a qual define pontuações para cada caixa de acordo com a presença da classe do objeto detectado.

Os possíveis locais de interesse são extraídos da imagem original, como demonstrado na Figura 6, e enviados para o bloco de detecção de Logotipo.

Figura 6 - Fluxo de execução da CNN de classificação de uniforme

4.2 VERIFICAÇÃO DO LOGOTIPO

Para evitar falsos positivos no reconhecimento de uniformes, que indicariam falhas no monitoramento, escolheu-se a localização de um objeto padrão que está obrigatoriamente presente em todos os modelos de uniformes: o logotipo da Instituição. Portanto na segunda etapa do método proposto, após o reconhecimento do uniforme, detecta-se o logotipo no uniforme.

Ao final do processo de reconhecimento do uniforme e do logotipo nele, obtém-se uma pontuação de confiança de classificação, a qual é posteriormente analisada para liberar o acesso à instituição. Caso a pessoa não esteja uniformizada ou o logotipo da instituição não seja localizado, o acesso é recusado.

O processo completo realizado pelo sistema de identificação de alunos uniformizados é apresentado na Figura 7 na forma de pseudocódigo.

Figura 7 - Pseudocódigo de Rotina para Detecção de Aluno Uniformizado

```

1: function detect_uniform()
2:     image = take_picture()
3:     clothes = detect_clothes(image)
4:     if clothes:
5:         logo = detect_logo(clothes)
6:         if logo:
7:             allow_access()
8:         else:
9:             deny_access()
10:    else:
11:        deny_access()

```

A função *detect_uniform()* realiza o processamento da imagem adquirida pela câmera de vídeo quando um aluno passa seu cartão de acesso na catraca, de acordo com o ilustrado na Figura 4. A imagem capturada é utilizada como entrada da Rede Neural Convolutacional que localiza o uniforme do aluno na imagem, por meio na função *detect_clothes()*. Caso seja o resultado da avaliação seja positivo, a saída da CNN é analisada pela segunda Rede Neural. A função *detect_logo()* recebe como entrada apenas a parte detectada anteriormente, ou seja, apenas o uniforme do aluno para localizar o logotipo nesta parte selecionada.

Por fim, somente se for detectado o uniforme o acesso é liberado, através da função *allow_access()*, caso contrário a função *deny_access()* não autoriza o acesso a instituição e aciona um alarme.

5. AVALIAÇÃO E RESULTADOS

A fim de se verificar a efetividade e a qualidade dos resultados fornecidos pelo sistema proposto, foram realizadas as avaliações de desempenho preliminares. Esta seção apresenta os detalhes das avaliações e os principais resultados obtidos.

5.1 CONFIGURAÇÕES

Os treinamentos das Redes Neurais foram realizados em um computador com o Sistema Operacional Ubuntu 18.04, com um Processador Intel i7 4770 contendo quatro núcleos e oito *threads* a 3.4GHz, 24GB de Memória RAM DDR3 1600MHz, uma placa de vídeo nVidia GeForce GTX1050Ti com 4GB GDDR5 e um SSD 240GB como memória secundária. Foi escolhida uma Rede Neural Convolucional sequencial com 28 camadas convolucionais seguidas da função de ativação *ReLU6* [31].

Optou-se por utilizar esta CNN devido à sua velocidade de processamento, o que é apropriado para o uso no microcomputador *Raspberry Pi 3*, que dispõe de baixo poder de processamento. A Rede Neural escolhida substitui as camadas convolucionais pelas chamadas convoluções separáveis em profundidade, as quais permitem a construção de uma rede rápida e precisa, pois, a fatoração realizada por estas convoluções reduzem drasticamente a computação e o tamanho do modelo [31]. Para o treinamento foram utilizadas 66 imagens de alunos com o uniforme do Instituto Federal de Mato Grosso, conforme amostra apresentada na Figura 8, e para as avaliações utilizou-se outras 66 imagens capturadas diretamente da câmera descrita acima.

Figura 8 - Conjunto de imagens utilizadas para treinar as Redes Neurais

5.2 TREINAMENTO DAS REDES NEURAIIS

A Rede Neural Convolutacional utilizada foi treinada com a base de dados (*data set*) gerada pelos autores, de modo a reconhecer o uniforme da instituição de ensino utilizada. Conforme mencionado acima, a fase de treinamento ocorre em duas etapas. Na primeira etapa, a Rede Neural Convolutacional é treinada para reconhecer o uniforme como um todo e posteriormente o treinamento visa identificar o logotipo da instituição contido no uniforme. Em cada um dos treinamentos, o experimento contou com 223.000 (duzentos e vinte e três mil) iterações, número suficiente para a convergência do treinamento. Dessa forma, os resultados dessas duas etapas do treinamento são apresentados abaixo.

A Figura 9 mostra a variação do erro de classificação durante o treinamento das Redes Neurais, referentes à detecção do uniforme e do logotipo. Observa-se que ambas as redes neurais começam a convergir por volta de 7.000 (sete mil) iterações. No entanto, nota-se que no treinamento da CNN relacionada ao logotipo o erro apresenta menor estabilidade. Observa-se também que o erro da CNN que identifica o uniforme decresce mais rapidamente do que o erro das demais. Isso ocorre provavelmente devido à complexidade da imagem do logotipo que possui mais detalhes. Por fim, a Figura 9 mostra também que por volta de 223.000 (duzentos e vinte e três mil) iterações o erro para ambas as CNNs tende a estabilizar.

Figura 9 - Erro de Classificação das CNN: uniforme, logotipo e logotipo com Swish

Com o propósito de melhorar o desempenho da CNN proposta, sobretudo em termos da precisão obtida nas avaliações abaixo, avaliou-se também o desempenho da Rede Neural com a função de ativação *Swish*. Esta alteração foi proposta com base nos resultados em [31], segundo o qual, a função *Swish* tende a produzir resultados mais precisos que a função *ReLU6* em modelos convolucionais profundos. Isso foi feito apenas para o treinamento de reconhecimento do logotipo, uma vez que este apresentou maior tempo de convergência, estável, em relação à CNN de classificação de uniformes. Os resultados da Figura 9 demonstram que a troca da função de ativação não impactou a convergência do erro na fase do treinamento. O treinamento das CNNs levou aproximadamente 3 dias para atingir 223 mil iterações

5.3 AVALIAÇÃO DAS REDES NEURAIS

Para avaliar a precisão do sistema proposto, considerando o encadeamento das duas CNNs, e determinar uma distância segura de uso e condições de iluminação adequadas, o sistema foi testado utilizando-se uma câmera de 8 *MegaPixels* com resolução de 1280x720 *pixels*, conectada ao *Raspberry Pi*. Conforme relatado acima, foram utilizadas nestas avaliações 66 imagens capturadas pela câmera, das quais 39 foram empregadas na avaliação de iluminação adequada e 27 na avaliação de distância da câmera.

Embora durante a fase de treinamento as funções de ativação *ReLU6* e *Swish* não tenham apresentado diferenças significativas, no que se refere à convergência do erro de saída das CNNs, ambas foram avaliadas aqui.

a) Precisão da detecção em função da distância da câmera:

A fim de encontrar a distância ideal para o posicionamento da câmera, foi avaliada a precisão conforme o indivíduo uniformizado se distancia da fonte de vídeo. O Figura 10 apresenta os resultados desta avaliação para as duas funções de ativação.

A função de ativação *ReLU6* apresentou melhor desempenho, com precisão máxima a uma distância de até cerca de 80 cm (oitenta centímetros), o que atende aos requisitos do modelo proposto como apresentado anteriormente na ilustração da Figura 4. Isso demonstra que apesar da *ReLU6* ser equivalente à *Swish* no treinamento, nesta avaliação, que considera a saída encadeada das CNNs, a *ReLU6* mostrou-se mais adequada.

Figura 10 - Avaliação do impacto da distância na precisão de classificação do sistema

b) Precisão de detecção em função da luminosidade:

Nestas avaliações foram comparadas 39 imagens com diferentes níveis de luminosidades aplicadas a elas a fim de verificar a influência deste parâmetro na precisão do sistema. Os resultados são apresentados na Figura 11. A partir dos resultados obtidos nas avaliações da subseção a), onde foi obtida precisão de 100% (cem por cento) para distâncias de até 60 cm (sessenta centímetros), as imagens analisadas nesta avaliação foram capturadas com distância entre 30 cm (trinta centímetros) e 60 (sessenta centímetros).

Nestes experimentos, a função de ativação *Swish* mostrou-se superior à função *ReLU6* em ambientes sob condições normais de luminosidade e sob forte iluminação. Por outro lado, os resultados também demonstram que sob níveis de alta luminosidade a precisão do sistema reduz acentuadamente, independente da função de ativação escolhida.

Figura 11 - Avaliação de do impacto da distância na precisão de classificação do sistema

5.4 DISCUSSÕES

Os resultados permitem afirmar que a escolha da função de ativação pode impactar a eficiência das CNNs utilizadas. No entanto, de acordo com o que foi obtido nas avaliações de distância da câmera e da intensidade de luminosidades, em que houve alternância de desempenho entre as funções *ReLU6* e *Swish*, é prudente dizer que a eficácia dessas funções depende dos cenários em que são aplicadas.

Nesta aplicação, por exemplo, em que a distância entre o aluno e a câmera será de no máximo 60 cm (sessenta centímetros) e a intensidade luminosa varia ao longo do dia, a função *Swish* é recomendada.

6. CONCLUSÕES

Este trabalho apresentou uma proposta de sistema de identificação de alunos uniformizados baseada em visão computacional. O Sistema funciona num microcomputador *Raspberry Pi*, usa redes neurais convolucionais, e é *open source*.

O sistema utilizou duas redes neurais encadeadas para detectar pessoas uniformizadas e o logotipo da instituição nos uniformes, como uma forma de fortalecer a precisão da identificação. Essas redes foram treinadas com imagens reais de estudantes uniformizados e avaliadas com imagens capturadas diretamente da câmera de vídeo conectada ao *Raspberry Pi*. A fim de otimizar o sistema para o cenário de aplicação, duas funções de ativação, *ReLU6* e *Swish*, foram avaliadas. Os resultados preliminares foram satisfatórios, uma vez que apesar da Rede Neural ter sido treinada com um pequeno conjunto

de dados, foi possível detectar corretamente alunos uniformizados e distingui-los daqueles que não portavam uniforme. É importante salientar que foi utilizado um hardware de pequeno poder computacional, o que demonstra que o sistema proposto pode ser facilmente embarcado em equipamento real, mantendo boa precisão e baixo custo. Neste trabalho, a função de ativação *Swish* apresentou ligeira vantagem sobre a *ReLU6*, considerando a distância da câmera e as condições de iluminação do ambiente de aplicação.

Como trabalho futuro, sugere-se o uso de segmentação de classes para o reconhecimento de uniformes, expandir a base de dados de treinamento e de testes e avaliar outros modelos de redes neurais convolucionais.

REFERÊNCIAS

- [1] S. Robers, A. Zhang, R. E. Morgan, and L. Musu-Gillette, "Indicators of School Crime and Safety : 2014," p. 90, 2015.
- [2] S. A. Rodrigues, "' Plano de Segurança para as escolas ' do estado de são paulo : o Controle das Discussões . Observam-se os resultados disso no uso de mecanismos privados por parte dos," pp. 333–348, 2002.
- [3] D. de Milano and L. B. Honorato, "VISÃO COMPUTACIONAL." UNICAMP Universidade Estadual de Campinas FT Faculdade de Tecnologia., 2014.
- [4] T. S. Tomasel, "APLICATIVO DE RECONHECIMENTO DE IMAGENS EM DISPOSITIVOS MÓVEIS PARA AMBIENTES PREVIAMENTE MAPEADOS." pp. 1–41, 2014.
- [5] T. Plötz, "Deep Learning for Human Activity Recognition in Mobile Computing," 2018.
- [6] X. Gao, J. Zhang, and Z. Wei, "Deep Learning for Sequence Pattern Recognition."
- [7] Y. Seo, "Image Classification of Fine-grained Fashion Image Based on Style Using Pre-Trained Convolutional Neural Network," pp. 387–390, 2018.
- [8] V. E. R. da Silva Filho, T. da S. Cavalcante, and E. Calvalcanti Neto, "Facial Biometrics for Access Control Using LBP Texture Measures and ELM Neural Networks," An. do XIII Work. Visão Comput., pp. 13–18, 2017.
- [9] D. V. Minatto, "Análise De Características Em Redes Neurais Artificiais Aplicadas Em Visão Computacional Com Ênfase Em Reconhecimento Facial," no. 1997, p. 97, 2016.
- [10] J. Zeng, J. Zeng, and X. Qiu, "Deep learning based forensic face verification in videos," 2017.
- [11] X. Han and Q. Du, "Research on Face Recognition Based on Deep Learning," pp. 53–58.
- [12] L. C. de P. Silva and F. S. Osório, "SISTEMA AUTÔNOMO E INTELIGENTE DE RECONHECIMENTO FACIAL RESTRITOS," pp. 1–6.
- [13] T. Ahmad, H. Studiawan, and T. T. Ramadhan, "Developing a Raspberry Pi-based Monitoring System for Detecting and Securing an Object," Int. Electron. Symp., pp. 125–129, 2014.
- [14] N. Patil, S. Ambatkar, and S. Kakde, "IoT Based Smart Surveillance Security System using Raspberry Pi," pp. 344–348, 2017.
- [15] W. G. De Souza, I. E. Fonseca, and F. A. Belo, "Modelo estimativo do fluxo de embarque e desembarque de passageiros em sistemas de monitoramento metroferroviário com base em técnicas de visão computacional," pp. 3–6, 2017.
- [16] Z. Hu, H. Yan, and X. Lin, "Clothing segmentation using foreground and background estimation based on the constrained Delaunay triangulation," Pattern Recognit., vol. 41, no. 5, pp. 1598–1609, 2008.

- [17] M. Yang and K. Yu, "Real-time clothing recognition in surveillance videos," Proc. - Int. Conf. Image Process. ICIP, pp. 2937–2940, 2011.
- [18] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," 2015.
- [19] S. HAYKIN, Redes neurais: princípios e prática. 2007.
- [20] Z. L. KOVÁCS, Redes neurais artificiais. 2002.
- [21] E. Bezerra, "Introdução à Aprendizagem Profunda," no. October, 2016.
- [22] G. Juraszek and S. Educacional, "Image Recognition of Products Using Bag of Visual Words and Convolutional Neural Networks," no. September, 2015.
- [23] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks An Introduction to Convolutional Neural Networks," no. November, 2015.
- [24] A. Santos, K. Aires, R. Veras, V. Uchôa, and L. Santos, "Uma Abordagem de Classificação de Imagens Dermatoscópicas Utilizando Aprendizado Profundo com Redes Neurais Convolucionais."
- [25] M. D. Zeiler and R. Fergus, "Stochastic Pooling for Regularization of Deep Convolutional Neural Networks," pp. 1–9, 2013.
- [26] S. Lazebnik, "Deep Convolutional Neural Networks for Image Classification," 2018 Int. Conf. Adv. Syst. Electr. Technol., pp. 1–10, 2014.
- [27] Moacir A. Ponti and G. B. P. da Costa, Como funciona o Deep Learning. 2017.
- [28] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, "Learning Activation Functions to Improve Deep Neural Networks," no. 2013, pp. 1–9, 2014.
- [29] P. Ramachandran, B. Zoph, and Q. V. Le, "SWISH: A SELF-GATED ACTIVATION FUNCTION," no. 1, pp. 1–12, 2017.
- [30] W. Liu et al., "SSD: Single shot multibox detector," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9905 LNCS, pp. 21–37, 2016.
- [31] A. G. Howard and W. Wang, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2012.