EAComp
Encontro Acadêmico de Computação

VI Encontro Acadêmico de Computação - UFMA
 ISSN: 2594-486X

Classificação de Faces de Personagens de Mangá

Ivan de Jesus Pereira Pinto[1], **Jessica Paloma Sousa Cardoso**[1]

[1]Universidade Federal do Maranhão
Caixa Postal 322 – 65065-545 – São Luís, MA – Brazil

navi1921@gmail.com, jessicacardosops@gmail.com

Abstract. *Japanese comics(Manga) have been a cultural phenomenon on audiences rating from young to mature. It has challenging aspects for recognizing,detecting and classifying the many elements on a manga page. In this work we evaluate one established technique to character face extraction, the HOG descriptors, against the modern choice of transfer learning through CNN, both being feed to SVMs classifiers. We found that CNN trained on human faces are able to generalize on better manga faces features than the traditional HOG.*

Resumo. *Histórias em quadrinhos japonesas têm sido um fenômeno cultural em audiências jovens e adultas. Também possui aspectos desafiadores para reconhecer, detectar e classificar os muitos elementos presentes em uma página de mangá. Nesse trabalho nós avaliamos uma técnica estabelecida para extração de faces, os descritores HOG, contra a escolha moderna de se utilizar CNN para extração, ambos sendo utilizados junto a SVMs classificadores. Nós descobrimos que o CNN treinados em faces humanas são capazes de generalizar melhor em faces de mangá do que o HOG tradicional.*

1. Introdução

A indústria de vendas de mangás/hqs digitais é um fenômeno recente com crescimento considerável na atualidade. Segundo [AnimesNetwork 2016] estima-se que as vendas combinada de mangás digitais e impressos foram cerca de 3.91 bi de dólares. O mangá digital por sua vez experimenta um crescimento recente de 27.5% do último ano.

Mangás são um tipo de história em quadrinhos com tonalidade preto e branco, com ordem de leitura da direita para a esquerda. Possui uma gama de elementos, sendo os mais comuns as onomatopeias, balões de fala contendo texto, personagens, e cenários de fundo, todos geralmente contidos em painéis ou quadros.

O uso de mangás como objeto de pesquisa científica é algo bem presente nos últimos anos. Métodos com altas precisões foram desenvolvidos para extração de painéis das páginas por [Arai and Herman 2010],[Pang et al. 2014]. Os balões de fala também já foram explorados tanto com o objetivo de sua detecção e classificação [Tanaka et al. 2010], quanto pelo propósito de extração de texto [Arai and Tolle 2011],[Rigaud et al. 2013].

Trabalhos voltados para detecção de personagens são em menor quantidade. [Sun et al. 2013] utilizou SIFT para detecção e identificação de personagens em comics, experimentados e validados em 6 títulos diferentes. [Yanagisawa et al. 2014] estudou o uso do HOG junto com SVM para detecção de face, utilizado 28 regiões de faces presente

em 9 páginas de mangás como treinamento. Em outro estudo utilizou a mesma técnica para detecção de olhos [Ishii et al. 2012] .

O nosso trabalho se diferencia dos demais ao ser treinado e validado em uma extensa base de mangás, que variam em gênero e período de publicação, e tem como objetivo a comparação ainda não estudada entre métodos de extração de características tradicionais como HOG e aprendizado por transferência com CNN, sendo aplicados para propósito de classificação por meio de SVMs. Extensões como detecção e reconhecimento de personagens serão abordadas em trabalhos futuros.

2. Metodologia

A metodologia seguida nesse trabalho aborda desde a segmentação das images que vão compor a base de dados(positivas e negativas), ao uso de técnicas específicas para extração de características como HOG e CNNs, que serão utilizadas para treinamento e predição com SVMs. O procedimento é ilustrado na Figura 1.

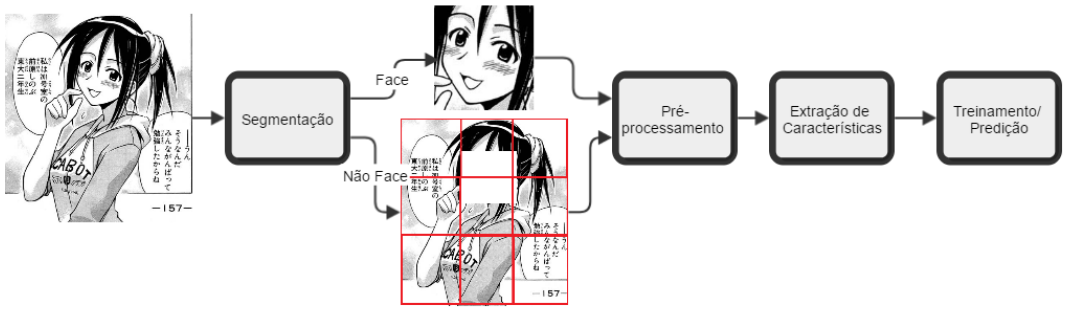

Figura 1. Metodologia

2.1. Segmentação da Base de Dados e Pré-Processamento

Enquanto é relativamente fácil se encontrar uma base de treinamento e validação para imagens como faces humanas, não se pode dizer o mesmo de imagens de mangá. No entanto, o trabalho de [Matsui et al. 2016] chamado Manga109 montou uma base de dados com 109 títulos dos mais variados mangás entre 1970 e 2010 marcados de acordo com seus públicos-alvos e gêneros. Consideramos essa base um conjunto apto para testarmos as técnicas propostas nesse trabalho.

Coletamos do Manga109 um total de 1200 imagens de faces dos personagens, ditas amostras positivas. Essa quantidade é ampliada artificialmente ao se aplicar rotação a cada 90 graus, resultando em 4800 faces. Como amostras negativas geramos 5000 imagens, retiradas das várias páginas de diferentes mangás do Manga109. Essas páginas são asseguradas de não conter faces, pela ausência delas ou sendo retiradas manualmente.

O processo de segmentação das faces possue ainda as seguintes particularidades :

1. A região extraída é quadrada, com ênfase na parte facial do personagem a fim de tentar minimizar elementos desnecessários do *background*
2. As faces possuem um tamanho mínimo de 64x64, algumas com oclusão

3. Imagens negativas foram geradas a partir de páginas de mangá, sendo que aquelas aproximadamente homogêneas(de 1 tonalidade somente) são descartadas por não conter informações relevantes

As faces segmentadas do Manga109 tem expressões que variam entre o mesmo título, e mais ainda entre diferentes títulos. A Figura 2. nos mostra a diferença entre uma imagem com elementos padrões a serem aprendidos: dois olhos,boca e nariz. Em contraste temos a imagem com esses elementos distorcidos ou exacerbados.

Figura 2. Variação entre faces

2.2. Extração de Características

Imagens de mangás possuem características diferente de imagens naturais, como a presença de linhas e tracejados descrevendo os personagens e outros elementos visuais. Desse modo, é desejável descritores que consigam extrair informações de linhas e bordas. Um candidato natural para esse trabalho é o HOG, que já foi abordado por [Yanagisawa et al. 2014].

Em contra partida o campo do aprendizado de máquina tem experimentado um recente crescimento do uso de modelos de várias camadas, ou deep. Em especial nas áreas de reconhecimento de padrões e visão computacional, a utilização de redes convolucionais profundas(CNN) vem ultrapassando as técnicas mais tradicionais. Exploramos nesse trabalho a abordagem baseada na utilização de CNN para extração de características, que obteve bons desempenhos em várias aplicações de classificações de objetos[Sharif Razavian et al. 2014].

2.2.1. HOG

A forte presença de bordas nas imagens de mangá faz do HOG um descritor desejável para extração de características. O algoritmo é ilustrado na Figura 3 e pode ser dividida nas seguintes etapas:

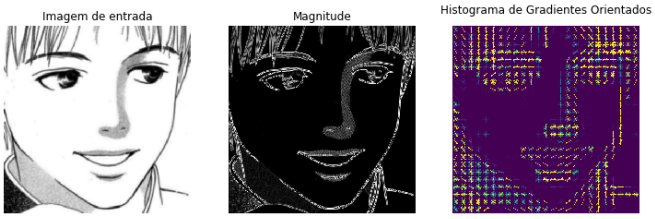

Figura 3. HOG

1. Pré-processamento da imagem de entrada, fazendo um resize de 64 x 64 .
2. Conversão em escala de cinza, normalização do contraste com uso de CLAHE [Zuiderveld 1994]
3. Cálculo dos gradientes horizontais e verticais, com suas magnitudes e direções. Informações não essenciais são descartadas, e contornos são destacados
4. A imagem é divida em células 6x6, com histogramas sendo calculados para cada cédula. Essa divisão é suficientemente grande para conter características como olhos/narizes,etc das faces que utilizamos. Obtém-se um histograma de gradientes correspondente a 9 direções discretas
5. Realiza-se a normalização dos histogramas em blocos de 2x2 células, dando robustez contra variações de luz
6. Concatena-se os histogramas em vetores por meio de uma janela deslizante, reunindo todos em um só vetor de características ao término.

2.2.2. Extração de Características com CNN

O uso de CNN vem sendo cada vez mais comum no reconhecimento de objetos. No entanto, o treinamento dessas redes profundas demanda tempo(dias a semanas) e custo(gpus), tornando difícil a aplicação para muitos casos. Recentemente [Sharif Razavian et al. 2014] descreveu a utilização de CNNs como forma de extração de características, e posterior treinamento em SVMs, obtendo índices competitivos com o estado da arte. Essa abordagem é também chamada de aprendizado por transferência.

A escolha de uma CNN adequada para o tipo de objeto que se deseja classificar é de suma importância, sendo importante considerar com o que elas foram pre-treinadas. Para esse trabalho utilizamos o VGG-Face, rede CNN baseada na rede profunda VGG-16 [Parkhi et al. 2015], que foi treinada em um dataset com cerca de 2622 identidades humanas. Sua arquitetura é mostrado na Figura 4, e descrita como:

1. Possui camadas de convolução 3x3, camadas de pooling 2x2, e 3 camadas totalmente conectadas.
2. Retira-se a camada softmax para predição, extraindo o descritor de 4096 dimensões da primeira camada FC.
3. O input deverá ser imagens 224x224x1 em escala de cinza

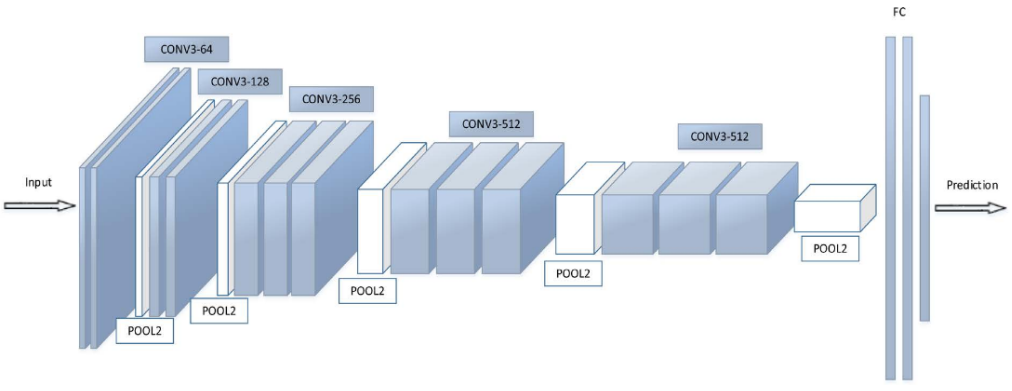

Figura 4. VGG-Face por [H. El Khiyari, H. Wechsler 2016]

3. Experimentos e Validação

Como explicado anteriormente, a base de dados contém 4800 faces positivas e 5000 imagens negativas, todas com proporção 1:1. Dividimos o conjunto de dados com uma proporção de 70% para treinamento e 30% para validação.

Ambas as características extraídas pelo HOG e CNN foram treinadas no classificador SVM linear, que é uma escolha inicial robusta e bastante utilizada para uma gama de problemas. O SVM tem seu hiper-parâmetro C com valor padrão 1.

Tabela 1				
Técnicas	Sensibilidade	Especificidade	Precisão	Acurácia
HOG+SVM	0.9456	0.9463	0.9157	0.94611
CNN+SVM	0.9673	0.9648	0.944	0.9658

Na tabela 1 vemos os resultados da validação nos dados de teste após o treinamento. Ambos conseguiram taxas altas nas estatísticas mostradas, com a maior diferença entre as abordagens sendo a taxa de precisão, onde o SVM com CNN é cerca de 3% melhor.

Um segundo experimento foi conduzido, utilizando-se de títulos de mangás que não estão presentes na base de dados. O objetivo é simular a performance dos SVMs em mangas desconhecidos, e quão bem as técnicas de extração conseguem generalizar para tais casos. Foram segmentadas 800 novas faces junto com 1000 novos negativos.

A tabela 2 nos mostra que o SVM treinado com HOG tem uma perda em todas as estatísticas, especialmente na sensibilidade, que indica o quanto de positivos foram acertados. No entanto o SVM com CNN se mostra robusto quanto as diferentes faces, melhorando as estatísticas anteriores.

Tabela 2				
Técnicas	Sensibilidade	Especificidade	Precisão	Acurácia
HOG+SVM	0.8533	0.8922	0.8729	0.8741
CNN+SVM	0.9661	0.9843	0.9816	0.9758

4. Conclusão

Nesse trabalho foi estudado abordagens para classificação de faces de personagens de mangá. Foi construído uma base de faces manualmente a partir do MangaDB para treinamento e validação. Experimentamos duas técnicas para extração de características e avaliamos sua performances junto ao classificador SVM. Concluímos que para a validação convencional, o HOG demonstrou ser uma boa escolha combinado com o SVM, e que uma CNN previamente treinada em faces humanas consegue ser generalizada para face nos mangas, gerando características ainda melhores que o HOG. Por fim verificamos que o SVM com HOG não mantém sua qualidade quando faces não incluídas na base são testadas, diferentemente do SVM com CNN que continua robusto. Em futuros trabalhos será explorada a utilização de outros classificadores como Random Forest, Redes Neurais, AdaBoost, e outros objetivos como expandir para detecção de personagens, realizar classificação de mangás por gênero e autor, dentre outros.

 ISSN: 2594-486X

Referências

AnimesNetwork (2016). Manga Sales Digital and Printed. https://goo.gl/OdV45x. [Online; acessado em 04-Junho-2017].

Arai, K. and Herman, T. (2010). Method for automatic e-comic scene frame extraction for reading comic on mobile devices. In *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on*, pages 370–375. IEEE.

Arai, K. and Tolle, H. (2011). Method for real time text extraction of digital manga comic. *International Journal of Image Processing (IJIP)*, 4(6):669–676.

H. El Khiyari, H. Wechsler (2016). Vgg-face cnn architecture. [Licensed by https://creativecommons.org/licenses/by/4.0/Online; accessed June 4, 2017].

Ishii, D., Yamazaki, T., and Watanabe, H. (2012). Multi size eye detection on digitized comic image. In *Proc. of IIEEJ 3rd Image Electronics and Visual Computing Workshop, 1P-4*.

Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K. (2016). Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, pages 1–28.

Pang, X., Cao, Y., Lau, R. W., and Chan, A. B. (2014). A robust panel extraction method for manga. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1125–1128. ACM.

Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In *BMVC*, volume 1, page 6.

Rigaud, C., Burie, J.-C., Ogier, J.-M., Karatzas, D., and Van de Weijer, J. (2013). An active contour model for speech balloon detection in comics. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 1240–1244. IEEE.

Sharif Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813.

Sun, W., Burie, J.-C., Ogier, J.-M., and Kise, K. (2013). Specific comic character detection using local feature matching. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 275–279. IEEE.

Tanaka, T., Toyama, F., Miyamichi, J., and Shoji, K. (2010). Detection and classification of speech balloons in comic images. *The Journal of the Institute of Image Information and Television*, 64(12):1933–1939.

Yanagisawa, H., Ishii, D., and Watanabe, H. (2014). Face detection for comic images with deformable part model. In *4th IIEEJ International Workshop on Image Electronics and Visual Computing (October 2014)*.

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc.