

# Identificação automática de pichação a partir de imagens urbanas

Eric K. Tokuda and Roberto M. Cesar-Jr.  
Institute of Mathematics and Statistics  
University of São Paulo (USP)  
Brazil

Claudio Silva  
Tandon School of Engineering  
New York University (NYU)  
USA

**Abstract**—Graffiti tagging is a common issue in great cities and local authorities are on the move to combat it. The tagging map of a city can be a useful tool as it may help to clean up highly saturated regions and discourage future acts in the neighbourhood and currently there is no way of getting a tagging map of a region in an automatic fashion and manual inspection or crowd participation are required. In this work, we describe a work in progress in creating an automatic way to get a tagging map of a city or region. It is based on the use of street view images and on the detection of graffiti tags in the images.

**Resumo**—A pichação é um problema comum em grandes cidades e autoridades locais se esforçam para combatê-la. O mapa de pichação de uma região pode ser um recurso muito útil, pois pode auxiliar no combate ao vandalismo em regiões com alto nível de pichações e também a limpeza de regiões saturadas para desestimular atos futuros na mesma região. Atualmente não existe uma maneira automática de se obter o mapa de pichação de uma região e atualmente ele é obtido pela inspeção manual da polícia ou pela participação popular. Nesse sentido, descrevemos um trabalho em andamento no qual propomos uma forma automática de obter um mapa de graffiti de uma região geográfica. Ele se baseia no uso de imagens com vista de rua e na detecção de pichação nas imagens.

## I. INTRODUÇÃO

Atualmente o grafite já faz parte do cenário das grandes cidades. Pode ser categorizado como grafite artístico ou pichação, como exemplificado na Figura 1, e enquanto o grafite é considerado uma expressão artística e, como tal, requer habilidades específicas, a pichação é geralmente um ato não autorizado que as pessoas simplesmente exibem frases ou nomes e a aceitação sobre pichação como arte é controversa [1], [2]. Em 2017, a prefeitura de São Paulo, Brasil decretou uma lei que <sup>1</sup> que definia uma multa de até R\$10.000,00 aos autores de crimes de pichação contra patrimônios públicos.

Atualmente não existe uma modo automático de se criar o mapa de pichação de uma região geográfica e a criação por inspeção manual é uma tarefa dispendiosa. Neste trabalho descrevemos o trabalho em curso de uma metodologia para a criação de um mapa de pichação baseado na segmentação de regiões pichadas em imagens urbanas de público acesso.

<sup>1</sup><https://www1.folha.uol.com.br/cotidiano/2017/02/1860352-doria-sanciona-lei-anti-pichacao-e-veta-ate-grafite-nao-autorizado.shtml>



Figura 1. Grafite e pichação. As pinturas de grafite são frutos de trabalhos meticulosos, enquanto as pichações são frequentemente atos não autorizados e compostos por letras e textos. Imagens obtidas a partir do serviço de visualização de ruas do [3].

## II. TRABALHOS RELACIONADOS

Trabalhos anteriores exploraram a tarefa de identificação de pichação em uma imagem [4]–[7] e paralelamente uma série de trabalhos utiliza dados geolocalizados para análises sociais, econômicas e culturais [8]–[10]. Nenhum trabalho, porém, investigou a variação da concentração de pichação nos diferentes bairros da cidade e sua correlação com indicadores socio-econômicos e culturais.

Ferramentas de combate à pichação [4]–[7] utilizam participação colaborativa e permitem que usuários equipados com telefones celulares reportem atos de pichação. Alguns trabalhos [11]–[13] atacam o problema de uma maneira diferente e tentam identificar o ato de pichar. Outros trabalhos recuperam pichações similares em um banco de dados de referência [14] usando componentes conexos e pontos-chave correspondentes em uma tentativa de associar pichação a gangues. Outra abordagem busca identificar a autoria do pichação [15] dada uma imagem de teste através da recuperação de imagens similares, cálculo de uma métrica baseada nos símbolos contidos, anotação manual e correspondência entre os pontos-chave das imagens de pichação e as da gangue.

Como um sinal da relevância do tema em nível global, a União Européia tem um projeto [16] dedicado à análise dos principais atores envolvidos nos atos de pichação, incluindo escritores, cidadãos, autoridades policiais e administração pública. Este projeto inclui entrevistas com as partes interessadas e o estabelecimento de uma plataforma web que permitem

a discussão e o compartilhamento de idéias sobre o tema de diferentes perspectivas. No Brasil, cidades altamente densas como São Paulo também enfrentam a presença generalizada de pichações na cidade [17].

A segmentação *semântica* é uma tarefa de Visão Computacional que visa dividir a imagem em classes conhecidas [18]. É uma tarefa complexa quando comparada com a classificação de imagens e a detecção de objetos, pois requer a classificação de cada pixel. A pesquisa nesta área é ativa e trabalhos recentes alcançam altos níveis de acurácia [18]–[22]. Uma tarefa relacionada é a segmentação *de instâncias*, onde o objetivo é adicionalmente identificar as instâncias. Por exemplo, no caso de dois objetos com intersecção, o método deve ser capaz de identificar os limites de duas instâncias. Trabalhos anteriores [23], [24] atacaram o problema precedendo o estágio de detecção de objetos com um estágio de segmentação. Trabalhos de aprendizagem profunda (*deep learning*) vêm atingindo os melhores resultados em diversas áreas da Visão Computacional [19], [25]–[27], o que se observa também na tarefa de segmentação. O trabalho de Mask-RCNN [28], aborda a tarefa de segmentação de instâncias executando as propostas de classificação e de segmentação de maneira paralela. [28] baseia-se na arquitetura Faster-RCNN [29], mas com uma ramificação adicional para segmentação de instância.

Serviços como [3], [30] fornecem acesso público a imagens com vista de rua. As imagens são obtidas em diferentes localizações geográficas, períodos de aquisição e pontos de vista. Uma série de trabalhos já utilizaram esse tipo de imagem para fins de análise urbana [31]–[33]. Os autores de [8] utilizam imagens de vista da rua para comparar os elementos arquitetônicos de diferentes regiões geográficas. O trabalho [34] propõe a avaliação da vegetação urbana usando o mesmo tipo de imagens.

### III. MATERIAIS E MÉTODOS

Para estimar com confiança o nível de pichação em uma região geográfica, propomos uma métrica, o *nível de pichação*, calculada utilizando a segmentação das áreas pichadas em imagens de vista de rua.

#### A. Aquisição

Uma região geográfica de interesse é inicialmente definida e as imagens dessa região são adquiridas. Idealmente, toda a região deve ser mapeada, mas devido a limitações da cobertura das imagens fornecidas e a restrições computacionais, apenas uma amostra é considerada. Existem diferentes maneiras de realizar a amostragem [35], podendo ser classificadas em amostragem *aleatória* e *sistemática*. A amostragem aleatória remove o viés da seleção, mas não garante cobertura, diferentemente da sistemática que garante cobertura com precisão arbitrária, mas adiciona viés.

Uma vez que a amostra geográfica é definida, consideram-se as vistas de rua em cada ponto amostrado sendo que idealmente toda a cena visível em torno do observador deve ser considerada. Uma única vista panorâmica de 360° pode ser usada, mas neste caso haverá distorções presentes nas

imagens obtidas. Alternativamente, pontos de vista complementares para cada local podem ser consideradas (veja a Figura 2).



Figura 2. Quatro visualizações da mesma localização geográfica. Imagens obtidas de [3].

#### B. Identificação de pichação

Dado o objetivo de quantificar o nível de pichação em um determinado local, uma maneira simples e direta seria classificar binariamente uma imagem se esta contém ou não pichação. No entanto, esta abordagem nos daria uma informação discreta e imprecisa de cada região e assim definimos o nível de pichação  $P(l)$  de uma localização geográfica  $l$  como a soma das áreas pichadas  $A$  em cada foto. Essa abordagem pode ser afetada pelo projeção e pela profundidade da cena. Assumimos que regiões distintas, dado uma amostra de tamanho mínimo, têm distribuição similar de projeções e de distância até os anteparos e, com essa suposição,  $P(l)$  pode ser usado para comparar regiões geográficas diferentes. Como representamos cada local  $l$  por um conjunto de  $k$  visões, definimos  $P(l)$  como a soma das áreas das regiões que contêm pichação em cada exibição. Podemos então agregar o nível de pichação por região geográfica e calcularmos a média dos níveis de pichação em nossa amostra de tamanho  $n$  (veja Equação 1).

$$P(R) = \frac{\sum_{j=1}^n \sum_{i \in l_j} A_i}{n} \quad (1)$$

Nós optamos pelo método Mask-RCNN [28] para nossa tarefa de segmentação, dado seu alto desempenho relatado em importantes benchmarks [36], [37]. Apesar de produzir segmentação e informação de instâncias, neste trabalho estamos apenas interessados na segmentação produzida.

Dada a inexistência de bases de dados de pichação disponíveis, criamos um conjunto de dados com imagens anotadas manualmente. Estes foram usados para treinar nosso modelo.

### IV. EXPERIMENTOS

Inicialmente, coletamos uma amostra piloto de 10.000 imagens da cidade e as regiões que contêm pichação foram identificadas manualmente. Nosso conjunto de treinamento é composto de 632 imagens anotadas manualmente. Usamos uma arquitetura de redes residuais de 101 camadas [38] e um modelo pré-treinado no conjunto de dados COCO [37]. Utilizamos uma taxa de aprendizado de 0,001 e um momento de 0,9 e treinamos por 80 épocas. Usamos o modelo obtido na iteração de número 30, dado o seu menor erro de validação (ver Figura 3). Utilizando a métrica de precisão média proposto em VOC 2007 [39], nosso modelo apresentou uma

precisão média de 0,57. A Figura 4 apresenta uma amostra das detecções avaliadas. O tempo para o processamento de uma única imagem foi de 0,69s em uma Geforce GTX 1050.

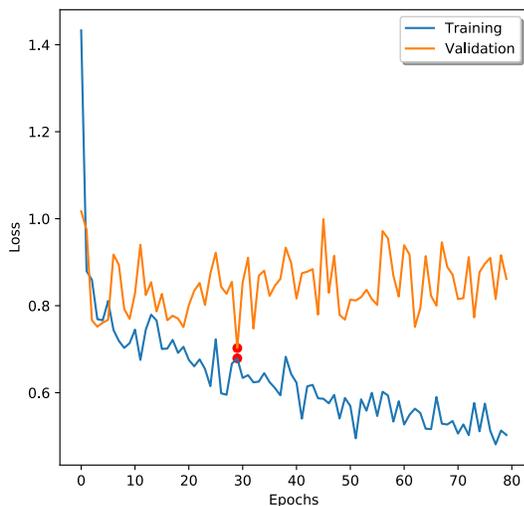


Figura 3. Perda do modelo durante a etapa de treinamento.

Na Figura 5 podemos ver a cobertura heterogênea das imagens fornecidas por [3]. Os dois distritos inferiores apresentaram pouca cobertura no momento de nossa aquisição, dada a natureza predominantemente rural e despovoadas dessas regiões e, portanto, não foram considerados neste estudo.

Usamos quatro visualizações para cada localização geográfica, espaçadas por 90°. Observe na Figura 2 como elementos de figuras adjacentes se cruzam, o que indica uma ampla cobertura de cada localização geográfica. 57% das imagens consideradas são de 2017, como pode ser visto na Tabela I.

Tabela I  
ANO DE AQUISIÇÃO DA AMOSTRA ANALISADA

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018
Pontos	1,241	16,311	207	422	2,182	4,563	4,211	39,391	317

Criamos uma malha sobre a extensão espacial da cidade com 134.624 pontos. Adotamos um espaçamento vertical e horizontal de 102 metros da nossa malha. Depois de eliminar pontos cujas imagens são de provedores externos e regiões não mapeadas (ver Figura 5), obtivemos uma cobertura geográfica de 68.752 pontos geográficos e 275.339 imagens no total.

## V. CONSIDERAÇÕES FINAIS

Este trabalho apresenta uma projeto em curso sobre o mapeamento automático de pichação em uma região geográfica. Utilizamos imagens de rua de uma região amostrada sistematicamente a partir da base do Google Maps [3] e identificamos as pichações em cada imagem. Propomos uma métrica para o nível de pichação de uma região geográfica.

Importante ressaltar que a métrica de pichação proposta é sensível à amostragem considerada, pois ela é calculada como uma média sobre os pontos *amostrados*. O método proposto também é sensível a escolha da abordagem de segmentação utilizada, cuja acurácia impacta diretamente o resultado final. Etapas em andamento incluem o teste com outros algoritmos de segmentação e uma análise de pichação na cidade de São Paulo. Passos futuros incluem a utilização do método com uma amostragem mais densa, comparação de diferentes regiões geográficas e identificação de pichações recentes através da utilização de bases espaço-temporais [40].

## AGRADECIMENTOS

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo, processos #2014/24918-0, #2015/22308-2 e ao CNPq, CAPES e NAP eScience - PRP - USP.

## REFERÊNCIAS

- [1] C. McAuliffe, “Graffiti or street art? negotiating the moral geographies of the creative city,” *Journal of urban affairs*, vol. 34, no. 2, pp. 189–206, 2012.
- [2] A. Young, *Street art, public city: Law, crime and the urban imagination*. Routledge, 2013.
- [3] Google, “Google Maps,” <https://www.google.com/maps>, 2005, [Last accessed April-2018].
- [4] Automated Regional Justice Information System (ARJIS), “Graffiti tracker,” <http://graffititracker.net/>, 2006, [Last accessed April-2018].
- [5] 594 Graffiti, LLC, “Tracking and Automated Graffiti Reporting System (TAGRS),” <http://www.594graffiti.com>, 2009, [Last accessed April-2018].
- [6] B. Archer, “Graffiti Tracking system,” <http://www.graffititrackingssystem.com/>, 2005, [Last accessed April-2018].
- [7] V. Ltd., “VandalTrack,” <https://www.vandaltrak.com.au/>, 2008, [Last accessed April-2018].
- [8] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. Efros, “What makes paris look like paris?” *ACM Transactions on Graphics*, vol. 31, no. 4, 2012.
- [9] B. Zhou, L. Liu, A. Oliva, and A. Torralba, “Recognizing city identity via attribute analysis of geo-tagged images,” in *European conference on computer vision*. Springer, 2014, pp. 519–534.
- [10] S. M. Arietta, A. A. Efros, R. Ramamoorthi, and M. Agrawala, “City forensics: Using visual elements to predict non-visual city attributes,” *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 2624–2633, 2014.
- [11] D. Angiati, G. Gera, S. Piva, and C. S. Regazzoni, “A novel method for graffiti detection using change detection algorithm,” in *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*. IEEE, 2005, pp. 242–246.
- [12] L. Di Stefano, F. Tombari, A. Lanza, S. Mattoccia, and S. Monti, “Graffiti detection using two views,” in *The Eighth International Workshop on Visual Surveillance-VIS2008*, 2008.
- [13] F. Tombari, L. Di Stefano, S. Mattoccia, and A. Zanetti, “Graffiti detection using a time-of-flight camera,” in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2008, pp. 645–654.
- [14] C. Yang, P. C. Wong, W. Ribarsky, and J. Fan, “Efficient graffiti image retrieval,” in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*. ACM, 2012, p. 36.
- [15] W. Tong, J.-E. Lee, R. Jin, and A. K. Jain, “Gang and moniker identification by graffiti matching,” in *Proceedings of the 3rd international ACM workshop on Multimedia in forensics and intelligence*. ACM, 2011, pp. 1–6.
- [16] S. Gmbh, “GRAFFOLUTION Awareness and Prevention Solutions against Graffiti Vandalism in Public Areas and Transport - Final report summary,” 2016.
- [17] The Guardian, “Pixação: the story behind São Paulo’s ‘angry’ alternative to graffiti,” <https://www.theguardian.com/cities/2016/jan/06/pixacao-the-story-behind-sao-paulos-angry-alternative-to-graffiti>, 2016, [Last accessed April-2018].



Figura 4. Amostra das detecções de pichação. Os valores descritos representam a confiança das detecções. Cenas provenientes de [3].



Figura 5. Cobertura por imagens com vista de rua por [3] sobre a cidade de São Paulo.

- [18] P. Arbeláez, B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik, "Semantic segmentation using regions and parts," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3378–3385.
- [19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [20] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.
- [21] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2017.
- [22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, 2015, pp. 3431–3440.
- [23] P. O. Pinheiro, R. Collobert, and P. Dollár, "Learning to segment object candidates," in *Advances in Neural Information Processing Systems*, 2015, pp. 1990–1998.
- [24] J. Dai, K. He, Y. Li, S. Ren, and J. Sun, "Instance-sensitive fully convolutional networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 534–549.
- [25] P. Hu and D. Ramanan, "Finding tiny faces," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017, pp. 1522–1530.
- [26] E. R. de Rezende, G. C. Ruppert, A. Theophilo, E. K. Tokuda, and T. Carvalho, "Exposing computer generated images by using deep convolutional neural networks," *Signal Processing: Image Communication*, 2018.
- [27] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot et al., "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, p. 484, 2016.
- [28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2980–2988.
- [29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.
- [30] Mapillary AB, "Mapillary," <https://www.mapillary.com/>, 2013, [Last accessed April-2018].
- [31] A. G. Rundle, M. D. Bader, C. A. Richards, K. M. Neckerman, and J. O. Teitler, "Using google street view to audit neighborhood environments," *American Journal of Preventive Medicine*, vol. 40, no. 1, pp. 94–100, 2011.
- [32] A. Torii, M. Havlena, and T. Pajdla, "From google street view to 3d city models," in *Computer vision workshops (ICCV Workshops), 2009 IEEE 12th international conference on*. IEEE, 2009, pp. 2188–2195.
- [33] E. K. Tokuda, G. B. A. Ferreira, C. Silva, and R. M. Cesar-Jr, "A novel semi-supervised detection approach with weak annotation," in *Image Analysis and Interpretation, 2018. SSIAP 2018. IEEE Southwest Symposium on*. IEEE, 2018.
- [34] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing street-level urban greenery using google street view and a modified green view index," *Urban Forestry & Urban Greening*, vol. 14, no. 3, pp. 675–685, 2015.
- [35] S. V. Stehman, "Basic probability sampling designs for thematic map accuracy assessment," *International Journal of remote sensing*, vol. 20, no. 12, pp. 2423–2441, 1999.
- [36] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.
- [37] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European Conference on Computer Vision*. Springer, 2014, pp. 740–755.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [39] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [40] E. K. Tokuda, Y. Lockerman, G. B. A. Ferreira, E. Sorrelgreen, D. Boyle, R. M. Cesar-Jr, and C. T. Silva, "A new approach for pedestrian density estimation using moving sensors and computer vision," *arXiv preprint arXiv*, 2018.