Metamodel development for annotated dataset capture in image detection machine-learning systems

Alex Cummaudo

December 12, 2017

Abstract

Abstract goes here

1 Introduction

In the last two decades, image detection systems have seen practical applications within various domains [1–8], but only recently has there been an apparent shift from heuristic-driven systems to machine-learning ones. Since the recent advancement of Convolutional Neural Networks (CNNs) from handwritten characters [9] to object detection [10] using the ImageNet dataset in 2012, there is increasing interest to improve the CNN [11–14] for such purposes. As it is typical for texture-based classifiers in image detection (such as CNNs) to require thousands of training samples [15, 16], providing *quality* annotations is therefore critical: significant flaws due to poor annotation directly affects Artificial Intelligence (AI) models trained with such data.

Not all training data is initially perfect, and therefore a cognitive aid that tracks data lineage helps us understand the flow of data to AI models: where is the data sourced from; how is it enriched, curated or updated; how does this affect the AI model; and how do we maintain a record of such changes? Without provenance tracking, tracing defects in future models is challenging. This paper proposes a methodology used to develop a novel annotation metamodel that offers a conceptual vocabulary and organising principle to improve the quality and efficiency of labelling training data.

References

[1] S. S. Tsai, D. M. Chen, V. Chandrasekhar, G. Takacs, N.-M. Cheung, R. Vedantham, R. Grzeszczuk, and B. Girod, "Mobile product recognition." *ACM Multimedia*, 2010.

[2] B. Girod, V. Chandrasekhar, D. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. Tsai, and R. Vedantham, "Mobile Visual Search," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 61–76, 2011.

[3] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W.-C. Chen, T. Bismpigiannis, R. Grzeszczuk, K. Pulli, and B. Girod, "Outdoors augmented reality on mobile phone using loxel-based visual feature organization." *Multimedia Information Retrieval*, 2008.

[4] J. Sivic and A. Zisserman, "Video Google - A Text Retrieval Approach to Object Matching in Videos." *ICCV*, 2003.

[5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[6] H. Bay, A. Ess, T. Tuytelaars, and L. J. Van Gool, "Speeded-Up Robust Features (SURF)." *Computer Vision and Image Understanding*, 2008.

[7] C. M. Jin, Z. Omar, and M. H. Jaward, "A mobile application of American sign language translation via image processing algorithms," in *2016 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2016, pp. 104–109.

[8] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. C. Jain, "Content-Based Image Retrieval at the End of the Early Years." *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks." *NIPS*, 2012.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014, pp. 580–587.

[12] R. B. Girshick, "Fast R-CNN." *ICCV*, 2015.

[13] S. Ren, K. He, R. B. Girshick, and J. S. 0001, "Faster R-CNN - Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.

[14] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN." *CoRR*, 2017.

[15] P. Shivakumara, T. Q. Phan, and C. L. Tan, "A Laplacian Approach to Multi-Oriented Text Detection in Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[16] D. Chen, J.-M. Odobez, and J.-P. Thiran, "A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods." *Sig. Proc. - Image Comm.*, 2004.