Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/64/20241368

Binning of biological image based on unsupervised semantic segmentation and no-reference image quality assessment

Feiyang Chen

Faculty of Science and Technology, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, 519000, China

beatrice_fy.chen@foxmail.com

Abstract. Accurate and efficient species classification is crucial for various applications, and machine-learning approaches have been widely employed for image-based taxonomic identification. However, challenges arise due to the variability in image quality and the need for extensive dataset preparation. In this study, we propose a model that combines unsupervised semantic segmentation by distilling feature correspondences (STEGO) and a no-reference image quality assessment model to assess the quality of images for taxonomic identification. STEGO enables the counting of objects in an image, while the image quality assessment model evaluates subject articulation. The model is applied to a dataset of Bombus hypnorum images from the Global Biodiversity Information Facility (GBIF) after data cleaning. The effectiveness of the model is demonstrated through the segmentation and scoring of images, providing a valuable tool for image recognition and quality assessment in taxonomic studies.

Keywords: Species classification, machine learning, image-based taxonomic identification, unsupervised semantic segmentation, image quality assessment.

1. Introduction

The need for accurate and fast species classification has spawned much research into the use of machine learning for image-based taxonomics. The learning is using different models to capture features on huge numbers of images, which raises two questions. The first is the tremendous amount of work when preparing for the dataset preparation. The pictures which are corrupted, and invalided (such as sketches, and maps) need to be deleted after manually inspecting the whole dataset [1]. Another is the quality of the image will greatly affect the accuracy of the recognition results, subtle similarities in the background can cause errors in the results (figure 1) and the bigger differences in background, illumination, and position would obscure minor differences between species [2]. One of the methods to circumvent this problem is using pictures with a blank background and placing the object in a similar position (dorsal view) like the research done by Fujisawa in 2023[3]. The pictures they used are under a blank background and in a relatively high resolution (less one is about 5.9 megapixels, figure 2). At the same time, only the picture of the subject that needs to be identified could be reliably identified [4].

DOI: 10.54254/2755-2721/64/20241368

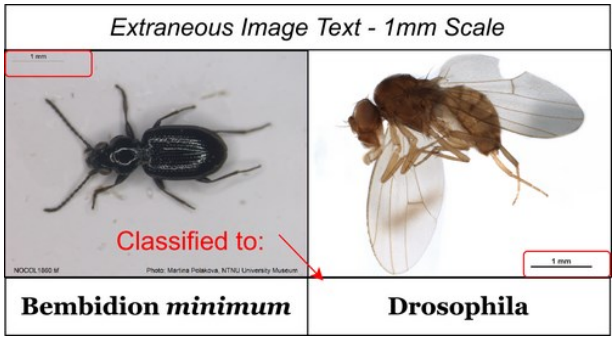

Figure 1. Misclassification due to irrelevant background information (1 mm scale line) shared by images across different species [5].

Figure 2. The sample picture of the camera (NIKON DS-Fi3 Microscope Camera) used by Fujisawa et al. in their local low-quality (LL) dataset on the website of Nikon [6].

2. Methodology

To improve the performance of machine learning or apply it to a broader context, the taxonomic systems should be trained under a larger and various content dataset. This brings the problem that the differences in the quality of images in the database will become larger, and the workload of data cleaning will increase.

A new model can be developed to solve this problem, which is to assign a value to the image. After considering various aspects, I decided to use two variables to represent the value, counting the Number of Objects and subject articulation. Counting the Number of Objects to describe the complexity of the environment of the creature in the picture, can be divided into different levels (figure 3). Subject articulation is to describe the clarity of the subject that needs to be identified, which can be represented by the image quality, and given a prediction mark (figure 4). The combination of two marks might generally define the final quality of the picture and be used in the quick filtering of databases.

Figure 3. Left: clear background, right: busy background [7].

DOI: 10.54254/2755-2721/64/20241368

Figure 4. Left: clear subject, right: blurred subject [7].

2.1. Counting the Number of Objects

Unsupervised semantic segmentation by distilling feature correspondences (STEGO) is an unsupervised semantic segmentation model based on the transformer model, developed by Hamilton et al., which can extract unsupervised features into high-quality discrete tags. The default mode of the model is trained on the CocoStuff dataset. [8] The output of the model can be used to count the number of objects in the picture by counting the number of areas with different tags.

For the objects that are not included in the tag list of training, the model can still make a good segmentation of the item. Some of the features which is hardly noticed by humans in the picture can be recognized by the segmentation and be a part of the quality assessment and evaluation of the results of follow-up experiments (figure 1). Compared to the other instance segmentation methods (like YOLOv5), semantic segmentation is done by STEGO on the whole picture, while YOLOv5 is focused on the main character of the picture [9]. STEGO is chosen to be used in this model since the judgment should be done on both the main character and background.

2.2. Subject Articulation

No-reference image quality assessment is a model that is based on the combination of CNN and transformer model based on the attention model. The model is developed to propose an approach to judge the picture objectively by extracting both local and non-local features of the image and then scoring the quality of the new image. A good generalization is a performance by the model and the robustness is strengthened by dong self-supervise to its self-consistency.[10]

This model is used to score the quality of the target area in the image. For subjects that are in motion, too small, out of focus, etc., this model will give a lower score; for that picture that clearly shows the details of the subject. Therefore, the score can be a way of judging to the subject articulation of the picture.

3. Realization

3.1. Database preparation

The pictures used in this study were from the Global Biodiversity Information Facility [11], which contains data not only collected by professional researchers but also from amateur observers. Therefore, the images downloaded from this database contain a variety of qualities and images in a variety of situations such as some invalid images (figure 5), damaged images (figure 6), and misfocused images (figure 7). The species of dataset selected in this study is Bombus hypnorum (Linnaeus, 1758). Having more than 94,095 georeferenced records, the bee is mainly found in Asia and Europe and has been recorded in North America, and GBIF provides 19,878 images in its database. Since the bee could be found under multiple postures in many places like buildings, the surface of the soil, or near plants, the images in the datasets have a variety of backgrounds. At the same time, the clarity of the bee is affected

DOI: 10.54254/2755-2721/64/20241368

by the state of motion of the bee and the photographer's level of shooting and equipment. For the above reasons, the dataset of Bombus hypnorum (Linnaeus, 1758) can be a suitable example for the study.

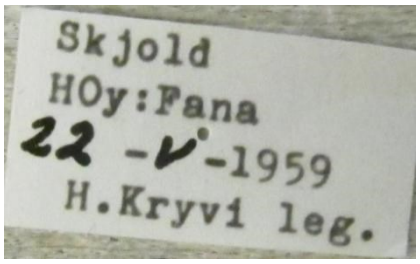

Figure 5. invalid image (no object) [7].

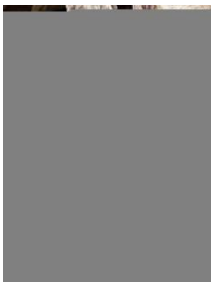

Figure 6. Damaged image [7].

Figure 7. Misfocused image (GBIF).

The data cleaning tool for the collected data is Easydata, an intelligent data service platform developed by Baidu [12]. The function of the platform that is mainly used in the cleaning is image deblurring. I used this function to filter low-definition images only to keep the images whose score of ambiguity scored by the platform is more than or equal to 75 (figure 8) to simulate a relatively strict image screening process, and it came out only about 27% of the images were preserved. However, there are still many low-quality images in the filtered dataset such as invalid images, images with blurred subjects and images with messy backgrounds were keep in the datasets. The next steps will work with 300 randomly selected images from the cleaned image set.

For the data that need to be put in the trained model of subject articulation, the divide of the area of the object is finished in the Easydata with its function of intelligent annotation, which can give the outcome of annotation of the area of the place of the object with only a few examples, and the correction is perfect. During this step, the picture that has no objects in it is deleted.

DOI: 10.54254/2755-2721/64/20241368

Figure 8. Pictures with different scores of ambiguity are provided by Baidu. From left to right: very blurred, blurred, clear, very clear [12].

3.2. Counting Number of Objects

The pre-trained model of STEGO is used for the semantic segmentation of the selected images. The type of region is counted according to whether there are similar classification label pixels adjacent to each other in eight directions.

Images with a more cluttered background have more areas. Counting results are often more than countable with the naked eye because some small, hard-to-see areas are segmented (such as undefined boundaries), creating a large number of segmented areas. The count of these small areas is retained because these small areas may also be recognized by the recognition program as certain features that affect the subsequent recognition model results.

3.3. Subject Articulation

The pre-trained model of No-reference image quality assessment was used to score the cropped images. As you can see from the results, blurred and clear images have ratings that can be distinguished.

Pictures with multiple bees may get a relatively low score when it comes to identifying image complexity. Therefore, to have a uniform basis for the final image quality classification, images with multiple bees will be rated 0 (lowest quality) at this stage.

3.4. Model integration

In the drawing of the coordinate system, the subject clarity is scored on the x-axis and the number of items is scored on the y-axis [fig 9]. For this coordinate system, the point represents the image, the lower the right point of the image for image recognition, the higher the quality of the image to the upper left, the lower the quality.

The statistical results of the data show that most of the data is concentrated in the subject resolution 20-60, the number of picture elements 0-1000. Some outliers may get bad results in the model that uses the clustered points as the main picture as the training data, so they can be eliminated when brought in.

DOI: 10.54254/2755-2721/64/20241368

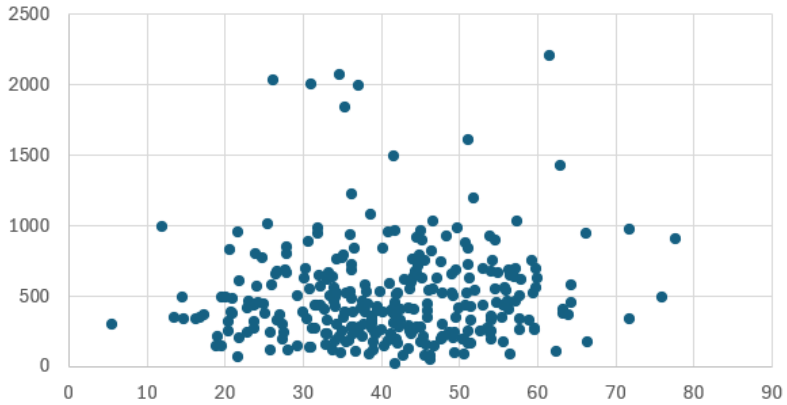

Figure 9. Result diagram.

4. Next Jobs to Do

The results of segmentation and scoring in this paper have some deviation from the actual image recognition criteria, which can be solved by finding a better model.

The model in this paper only uses the pre-training data of the corresponding model. To achieve better scoring or differentiation effect, the model should be trained with new data to make the model more targeted.

The model in this article uses two criteria, and multiple criteria can be added to complete a more complex model to make the evaluation more scientific and standard.

5. Conclusion

In this study, we developed a model that combines unsupervised semantic segmentation and image quality assessment for taxonomic identification. The model effectively counts the number of objects in an image using STEGO, while subject articulation is evaluated using a no-reference image quality assessment model. The model was applied to a dataset of Bombus hypnorum images, and the results showed its capability to distinguish image quality and complexity. However, further improvements can be made by refining the segmentation and quality assessment models and incorporating additional criteria for a more comprehensive evaluation. The proposed model provides a valuable approach for addressing the challenges associated with image-based taxonomic identification, offering potential applications in biodiversity research and conservation efforts.

References

[1] Zhou, Z., Fu, G., Fang, Y., Yuan, Y., Shen, H., Wang, C., Xu, X., Zhou, P., & Pan, X. (2023). EchoAI: A deep-learning-based model for classification of echinoderms in global oceans. Frontiers in Marine Science.

[2] Alexander Knyshov, Samantha Hoang, Christiane Weirauch, pre-trained Convolutional Neural Networks Perform Well in a Challenging Test Case: Identification of Plant Bugs (Hemiptera: Miridae) Using a Small Number of Training Images, Insect Systematics and Diversity, Volume 5, Issue 2, March 2021, 3, https://doi.org/10.1093/isd/ixab004

[3] Fujisawa, T., Noguerales, V., Meramveliotakis, E., Papadopoulou, A., & Vogler, A.P. (2023). Image ‐ based taxonomic classification of bulk insect biodiversity samples using deep learning and domain adaptation. Systematic Entomology, 48, 387 - 401.

[4] Xu, D., Zhao, Y., Hao, X., & Meng, X. (2023). Pink-Eggs Dataset V1: A Step Toward Invasive Species Management Using Deep Learning Embedded Solutions. ArXiv, abs/2305.09302.

[5] Badirli, S., Picard, C., Mohler, G.O., Richert, F., Akata, Z., & Dundar, M. (2023). Classifying the unknown: Insect identification with deep hierarchical Bayesian learning. Methods in Ecology and Evolution, 14, 1515 - 1530.

DOI: 10.54254/2755-2721/64/20241368

[6] Nikon. (2023). DS-Fi3 | Cameras | Microscope Products | Nikon Instruments Inc. (Nikon), from Nikon Instruments Inc.: https://www.microscope.healthcare.nikon.com/products/cameras/ds-fi3
[7] GBIF.org (23 August 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.s5tc7z
[8] Hamilton, M., Zhang, Z., Hariharan, B., Snavely, N., & Freeman, W.T. (2022). Unsupervised Semantic Segmentation by Distilling Feature Correspondences. ArXiv, abs/2203.08414.
[9] Forest Fire Detection and Notification Method Based on AI and IoT Approaches - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/YOLOv5-SOTA-Real-Time-Instance-Segmentation_fig2_367986616 [accessed 6 Feb 2024]
[10] Golestaneh, S. & Dadsetan, Saba & Kitani, Kris. (2022). No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. 3989-3999. 10.1109/WACV51458.2022.00404.
[11] What Is GBIF? (n.d.). What-Is-Gbif. https://www.gbif.org/what-is-gbif
[12] EasyData. (2023). Easy data. https://ai.baidu.com/easydata/