Solving the Camouflage Object Segmentation Problem with Multitask-Learning and Meta-Learning Augmented LSTM

William Cai; Xin Zhang

1 Objective

This project aims to 1) first apply both the multitask-learning technique and the meta-learning technique to train a recurrent neural network to identify camouflaged objects in different settings and then 2) compare the performance of these two techniques with the current state-of-the-art models that do camouflaged objects identification. To leverage meta-learning on a consolidated meta-camouflage dataset to enhance the learning for a selected deep neural network that will perform the task of Camouflage Object Segmentation (COS). COS is a subcategory of object detection in the field of computer vision and is about finding out the camouflaged objects that are "seamlessly" embedded in their surroundings [1]. COS plays an important role in many areas including medical imaging diagnosis (e.g. lung infection segmentation), detection of unqualified products on automatic production lines in the manufacturing industry, detection of personnel in search-and-rescue missions, and so on [1]. Although there are a number of datasets that can be used to train models to perform the task of COS, these datasets are generally small compared to large datasets like ImageNet thus limiting the performance of the trained models. And the objective of this project is to address this problem.

2 Related work

Research into camouflaged object segmentation has a long and rich history in many fields such as biology and art. There have been several networks like the Positioning and Focus Network (PFNet), content-aware pre-activated residual UNet (CAPA-ResUNet), and several deep learning-based approaches (e.g. deep network SINet) [2] [3] [4] that achieve promising performance to some extent. Nonetheless, when applying these trained models to various camouflage datasets, the performance of these models becomes inconsistent. Moreover, indexes – ranging from 0 to 1 – of these models that are used to measure performance rarely score more than 0.9 which is an indication of room for further improvement. A major issue for these previous attempts is that the datasets that are dedicated to camouflage object segmentation training are too small. If one combines all images from the current camouflage object datasets all together (CHAMELEON [5], CAMO [3], COD10K [2]), one will only have 6392 images in total compared to a total number of 1331167 images in ImageNet. Overall, the related works we have discussed fall short of achieving consistent and outstanding performance due to the limited datasets available. Thus, this project will set out to experiment with the multitask-learning technique and the meta-learning technique to see if we can solve the problems from the related work we have presented above.

3 Technical Outline

This project will use the datasets of CHAMELEON, CAMO, and COD10K. In addition, the project will use the MoCA dataset to expand the size of training data. Data from CHAMELEON, CAMO, COVID-19 CT Lung and Infection Segmentation Dataset, and COD10K come in the form of images and can be fed into the model training pipeline directly. On the other hand, data from the MoCA dataset will require some preprocessing before feeding into the pipeline since they are in video format.

After the data cleaning and processing, the project will involve the assembly of a 3-layer LSTM model as a vanilla model to test the basic functionality of the entire training pipeline and the evaluation process. We expect the vanilla model to have moderate performance on the task, but this preliminary attempt will enable us to give an estimation of the scale and complexity of the model that we wish to work on. The project will then implement the multitask-learning technique on top of the 3-layer LSTM model. Specifically, after the multitask-learning feature is added to the model, we wish to recover the performance of the naive model by degenerating the multitask into a single task. We will then experiment with parameter sharing on a variety of levels to see how this will impact the performance of the multitask-learning augmented LSTM model. Finally, the project will implement the meta-learning technique on top of another vanilla 3-layer LSTM model. We will experiment with different numbers of shots (k-shot learning) and ways (N-way classification) to see how this will impact the performance of the meta-learning augmented LSTM model.

To evaluate the effectiveness of these techniques on the performance of the model and to compare the performance with the ones from the other paper, we will need to compute four indexes and compare them with those presented in the other paper. The four indexes, Structure Measure, Enhanced-alignment Measure, Comprehensive evaluation index, and Mean Absolute Error, are established as industry standards in COS literature. Therefore we are able to gauge our model performance in a holistic metric.

Overall, the project will have the following novel technical contributions:

- Transforming the MoCA dataset from a video dataset into an image dataset that can be easily fed into a COSmodel training pipeline.
- Implementing the multitask-learning technique on top of a 3-layer LSTM model.
- Implementing the meta-learning technique on top of a 3-layer LSTM model.

4 Team Contributions

The given datasets(MoCA, CAMO, etc.) comprise images of identical resolution and format. Specifically, each sample in MoCA dataset is a video of 30k frames featuring a species of animal. Proper formatting and video editing are to be applied to samples of MoCA to standardize data (done by Xin Zhang). Furthermore, certain samples of MoCA have their labels missing. Therefore, it is necessary to perform manual labeling of images that identify the matrices where the camouflage objects are located (done by William Cai).

Under obligations of good faith, both of the members (Xin Zhang and William Cai) have performed necessary literature reviews in related works in COS and camouflage object dataset. Out of all the works reviewed so far, a model is to be selected as the benchmark model whose satisfactory performance the team is the most confident in. The implementation of the main model architecture will be done in the fashion of pair coding. Subsequent secondary works, such as parameter tuning, architecture twisting, and novel optimization shortcuts, will be split among members who will independently explore ideas in the given area of improvement. The team will convene on a regular basis to synchronize the efforts.

References

[1] Ji G. P. Wei Z. Yang X. Wei X. Fan D. P. Mei, H. Camouflaged object segmentation with distraction mining. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[2] Ji G. P. Sun G. Cheng M. M. Shen J. Shao L Fan, D. P. Camouflaged object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020.

[3] Nguyen T. V. Nie Z. Tran M. T. Sugimoto A. Le, T. N. Anabranch network for camouflaged object segmentation. In *Computer Vision and Image Understanding*, 2019.

[4] Le T. N. Nguyen K. D. Tran M. T. Do T. T. Nguyen T. V. Yan, J. Mirrornet: Bio-inspired camouflaged object segmentation. In *IEEE Access*, 2021.

[5] Abdulameer H. Błaszczyk J. Depta T. Kornacki A. Kozieł P. Skurowski, P. Animal camouflage analysis: Chameleon database. 2018.