Final Project Proposal
Deep Unsupervised Saliency Detection

João Victor da Silva Guerra, Leonardo Alves de Melo, and Marcos Felipe de Menezes Mota [*]

Saliency detection targets visually interesting regions in images that can attract human visual attention, which generates a namely saliency map that identifies pixel's unique features. These saliency maps aims at simplifying and/or changing the image into a more meaningful representation to analyze [1], [2]. In computer vision, salient object detection is commonly a process consist of two stages: (1) detect the most salient object; (2) segment the accurate object region [3]. In recent years, with the powerful learning capabilities of deep neural networks, convolutional models have been employed to address the task of salient object detection.

Deep learning saliency maps were introduced in [2], which applied visualisation techniques to compute images, saliency maps being one of them. Building upon the capabilities of convolutional neural networks (CNN), these deep supervised saliency detection methods achieved state-of-the-art performance with greater performance compared to unsupervised methods [1]. However, such supervised methods depends on large-scale manual supervision as pixel-level human annotations, which is highly time-consuming, labor-intensive and could hinder the generalization ability of the models [1], [3]. In the literature, "Supervision by Fusion" (SBF; [4]) have been the first successful unsupervised learning framework to train a salient object detector, which employs a two-stream framework to create supervisory signals through a intra-image and inter-image fusion processes. Aftwards, Deep Noise Model-based Saliency Detector (DNMSD; [1]) presented an end-to-end deep learning framework that learn saliency maps from multiple noisy unsupervised saliency methods. Instead of removing noise in saliency labeling from unsupervised saliency methods with fusion strategies as in [4], authors explicitly learn an adaptive noise from noisy saliency maps. The latter method outperforms traditional unsupervised method and also achieves similar performance with current state-of-the-art deep supervised saliency detection methods [1].
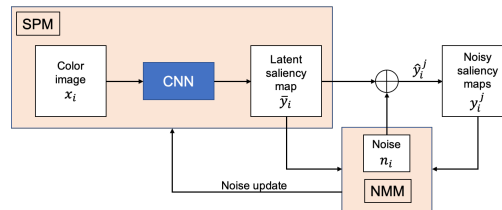

Figure 1. Schematic representation of the deep unsupervised detection framework.

In our project, we propose to reproduce the work of [1] with MSRA-B dataset [5], 5000 images with pixel-accurate object labeling, which most images only have one salient object. Further, the deep unsupervised detection framework (Fig. 1) consists of two connected modules: a saliency prediction model (SPM) that maps from a color image to a latent saliency map with a noise estimation, and a noise modelling module (NMM) that learns an adaptive noise in noisy saliency maps and updates the noise estimation with the saliency prediction and the noisy salient maps. Both modules works collaboratively with a joint optimization of their parameters.

References

[1] J. Zhang, T. Zhang, Y. Dai, M. Harandi, and R. I. Hartley, "Deep unsupervised saliency detection: A multiple noisy labeling perspective," *CoRR*, vol. abs/1803.10910, 2018. arXiv: 1803.10910. [Online]. Available: http://arxiv.org/abs/1803.10910.

[2] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *CoRR*, vol. abs/1312.6034, 2014. arXiv: 1312.6034. [Online]. Available: https://arxiv.org/pdf/1312.6034.pdf.

[*] 117410, 156188 and 211893. j117410@dac.unicamp.br, leonardo.alves.melo.1995@gmail.com, and marcos.mota@ic.unicamp.br

[3] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A survey," *CoRR*, vol. abs/1411.5878, 2014. arXiv: 1411.5878. [Online]. Available: http://arxiv.org/abs/1411.5878.

[4] D. Zhang, J. Han, and Y. Zhang, "Supervision by fusion: Towards unsupervised learning of deep salient object detector," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4068–4076.

[5] J. Wang, H. Jiang, Z. Yuan, M.-M. Cheng, X. Hu, and N. Zheng, "Salient object detection: A discriminative regional feature integration approach," *International Journal of Computer Vision*, vol. 123, no. 2, pp. 251–268, 2017, ISSN: 1573-1405. DOI: 10.1007/s11263-016-0977-3.