Going Beyond Feature Pyramids in Face Detection

Supervision: Furkan Kasım, Prof. Dr. Manuel Günther

Face detection is a fundamental task in computer vision, serving as the backbone for applications ranging from security systems to social media functionalities where identifying and localizing human faces within digital images is crucial. The evolution of deep learning, especially through convolutional neural networks (CNNs), has significantly enhanced the capabilities in this area. CNNs are skilled at analyzing spatial data, effectively learning and extracting hierarchical features from images that range from basic textures to complex objects like human faces.

In the realm of state-of-the-art face detection algorithms such as RetinaFace (Deng et al., 2020), RefineFace (Zhang et al., 2020), TinaFace (Zhu et al., 2020), and MogFace (Liu et al., 2021), a common architectural framework involving a backbone, neck, and head is employed. These systems leverage the strengths of CNNs to extract deep features (backbone), employ feature pyramid networks (FPN) (Lin et al., 2017) to handle various face scales (neck), and refine detection outputs at different network stages (head). In these systems, the complexity of handling face detections at multiple scales – ranging from very large to very small – still remains a significant challenge. Current methods like FPNs, while effective, increase computational complexity and can be resource-intensive.

This master thesis aims to streamline the existing architecture by replacing FPNs with pooling layers, possibly applying simple attention mechanisms. This approach uses different pooling sizes tailored to specific feature levels directly derived from the base network, simplifying the model while maintaining effectiveness. While larger pooling kernels are used for detecting larger faces, smaller ones are for smaller faces, aiming to enhance detection precision across all scales with reduced complexity. This method could provide a more efficient solution while still capturing the nuances necessary for high-quality face detection in diverse imaging conditions.

As a baseline model, RetinaFace will be used, replacing its FPN component with proposed pooling layers. The dataset chosen for training is WiderFace (Yang et al., 2016), which is renowned for its diversity in face scales and scenarios. For evaluation, the proposed model will be tested on the WiderFace validation and test subsets, with potential further evaluations on other common face detection benchmarks such as AFW (Zhu and Ramanan, 2012), PASCAL Face (Yan et al., 2014), FDDB (Jain and Learned-Miller, 2010). The primary evaluation metrics will include the Average Precision (AP) score, calculated from precision-recall curves, and an adaptation of the Free-response Receiver Operating Characteristic (FROC) curve, which plots the True Positive Detection Rate (TPDR) against the False Positive Detection Per Image (FPDPI) (Kasim et al., 2024). This analysis also provides an excellent opportunity to examine whether there is a correlation between these metrics. The expected outcomes of this thesis include a comprehensive understanding of current face detection systems, an investigation into the effectiveness of pooling layers in addressing multi-scale scenarios, and the development of a streamlined model architecture that could potentially set new benchmarks in face detection accuracy.

Requirements

- A reasonable understanding of deep neural networks.
- Successfully passing of the Deep Learning Course.
- Programming experience in Python and a deep learning framework, preferably PyTorch.

References

Deng, J., Guo, J., Ververas, E., Kotsia, I., and Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jain, V. and Learned-Miller, E. (2010). Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst.

Kasim, F., Boult, T. E., Mora, R., Biesseck, B., Ribeiro, R., Schlueter, J., Repak, T., Vareto, R. H., Menotti, D., Schwartz, W. R., and Günther, M. (2024). Watchlist challenge: 3rd open-set face detection and identification. In *International Joint Conference on Biometrics (IJCB)*.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Liu, Y., Wang, F., Sun, B., and Li, H. (2021). Mogface: Towards a deeper appreciation on face detection. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4083–4092.

Yan, J., Zhang, X., Lei, Z., and Li, S. Z. (2014). Face detection by structural models. *Image and Vision Computing*, 32(10):790–799.

Yang, S., Luo, P., Loy, C. C., and Tang, X. (2016). WIDER FACE: A face detection benchmark. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, S., Chi, C., Lei, Z., and Li, S. (2020). Refineface: Refinement neural network for high performance face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP:1–1.

Zhu, X. and Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2879–2886.

Zhu, Y., Cai, H., Zhang, S., Wang, C., and Xiong, Y. (2020). Tinaface: Strong but simple baseline for face detection. *arXiv preprint arXiv:2011.13183*.