Research Proposal

Lightweight Facial Landmark Detection using Multi-Teacher Knowledge Distillation

1. Research purposes

Facial Landmark Detection is a critical step in applications such as face recognition and facial expression analysis. The advancement of deep learning has significantly improved the accuracy of facial landmark detection. However, most state-of-the-art detection algorithms that achieve high accuracy rely on large backbone networks, such as Hourglass [1], ResNet [2], and HR-Net [3]. Consequently, there is a need for lightweight neural network architectures with smaller model sizes and lower computational costs to fulfill the requirements of fast and accurate facial landmark detection. Examples of such lightweight models include MobileNet [4] and ShuffleNet [5]. However, directly applying these lightweight models to facial alignment may severely compromise performance.

Therefore, my future research direction involves leveraging a Multi-Teacher Knowledge Distillation Network, combining channel-wise distillation and adversarial learning methods, to lighten the HR-Net network and achieve a lightweight facial landmark detection model with higher accuracy.

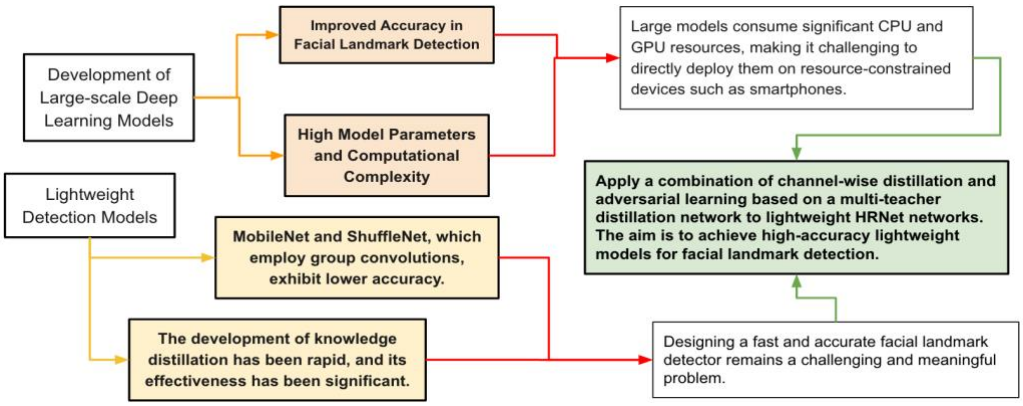

Figure 1 Reversal of research purpose

2. Literature Review

(1) Facial Landmark Detection

In recent years, deep learning-based frameworks [6], [7], [8] for facial landmark detection have shown significant advantages over traditional algorithms in terms of accuracy and efficiency. The heatmap regression method [8] generates heatmaps by producing Gaussian distributions on channels and predicts the points with the highest response on the heatmaps. Wang et al. [3] introduced the HRNet, a deep high-resolution representation learning network that employs a multi-branch parallel structure, enabling the network to preserve feature information at multiple resolutions simultaneously, thereby enhancing its perception of details.

While the heatmap regression method achieves better performance, it requires deeper networks and more parameters, resulting in computational complexity and slower processing speeds on devices with limited memory and computational resources (such as smartphones and robots).

(2) Knowledge Distillation

Knowledge Distillation is a popular model compression technique. Hinton et al. [9] proposed the concept of knowledge distillation in 2015 and applied it to image classification tasks. This technique suggests that the soft probabilities (soft targets) output by a trained teacher network contain more information about data points. Student networks can mimic these probabilities to absorb knowledge discovered by the teacher, beyond the information provided by training labels. In the following years, various variants of knowledge distillation have emerged.

Zhou et al. [10] introduced a technique called channel-wise knowledge distillation in 2021, which transfers the mean and variance of teacher model feature maps to the student model. This method focuses on information in the channel dimension, reducing the number of parameters and computations in the model, and helping the student model better learn the teacher model's feature representation in the channel dimension.

You et al. [11], in their 2017 paper "Learning from multiple teacher networks," proposed using multiple teacher networks to improve the performance of student networks. They introduced a loss function that encourages the student network to match the predictions of all teacher networks simultaneously, addressing limitations such as overfitting or bias in a single teacher network.

Overall, knowledge distillation has become a powerful tool for model compression, demonstrating effectiveness in various applications. In this research direction, we plan to apply the principles of knowledge distillation and its variants to better transfer knowledge from large models to more efficient small models.

(3)Adversarial Learning

The primary goal of Adversarial Learning is to enhance the robustness of machine learning algorithms by adversarial examples, enabling them to perform well in the face of unknown data. Goodfellow et al. [12] introduced Generative Adversarial Networks (GANs) in 2015, which train two neural networks, a generator and a discriminator. The generator is responsible for generating adversarial samples, while the discriminator attempts to distinguish between adversarial and real samples.

(4) Knowledge Distillation Approach for Facial Landmark Detection

Fard et al. [13] proposed a multi-teacher knowledge distillation approach for facial landmark detection in 2022. They utilized a new loss function and different soft-label heatmaps to train a lightweight student network for facial landmark detection. However, the authors also noted that while the performance of the new student network improved compared to MobileNet, there is still a certain gap in accuracy compared to more advanced models. This research project aims to draw inspiration from their approach and utilize new loss functions and more advanced models to distill a higher-accuracy lightweight facial landmark detection model.

3. Research Objectives

The aim of this research project is to apply knowledge distillation techniques to lightweight facial landmark detection models in order to achieve more efficient and accurate facial landmark detectors. The specific objectives of the future research direction include:

(1) Explore a multi-teacher distillation approach that combines channel knowledge distillation and adversarial learning techniques to lightweight the HRNet network for achieving high-accuracy lightweight facial landmark detection algorithms.

(2) Improve the accuracy and robustness of lightweight models by focusing on and integrating channel information.

(3) Investigate the application of adversarial learning in multi-teacher knowledge distillation to further enhance the model's generalization ability and resistance to interference.

(4) Evaluate the performance of the proposed methods and compare them with existing lightweight facial landmark detection models as well as other facial landmark detection models within this research direction.

4. Research Methods

This research project proposes a multi-teacher knowledge distillation framework composed o f two teacher networks and one student network:

1.Perform data preprocessing on the WFLW[14] and 300W[15] datasets.

2. Use High-Resolution Representations Network's HRNet-W32 and HRNetV2p models as the two teacher networks and reduce the number of channels in HRNet for the student network.

3. Calculate the feature map weight information of HRNet-W32 and the student network, and pas s the soft targets within the channels to the student network.

4. Employ adversarial learning by having the two teachers compare their output heatmaps with th e student network's output heatmap. This further enhances the model's generalization and resista nce to interference.

5. Compute the total loss function and perform backpropagation on the student network.

6. Evaluate the performance of the proposed method and compare it with existing lightweight faci al landmark detection models.

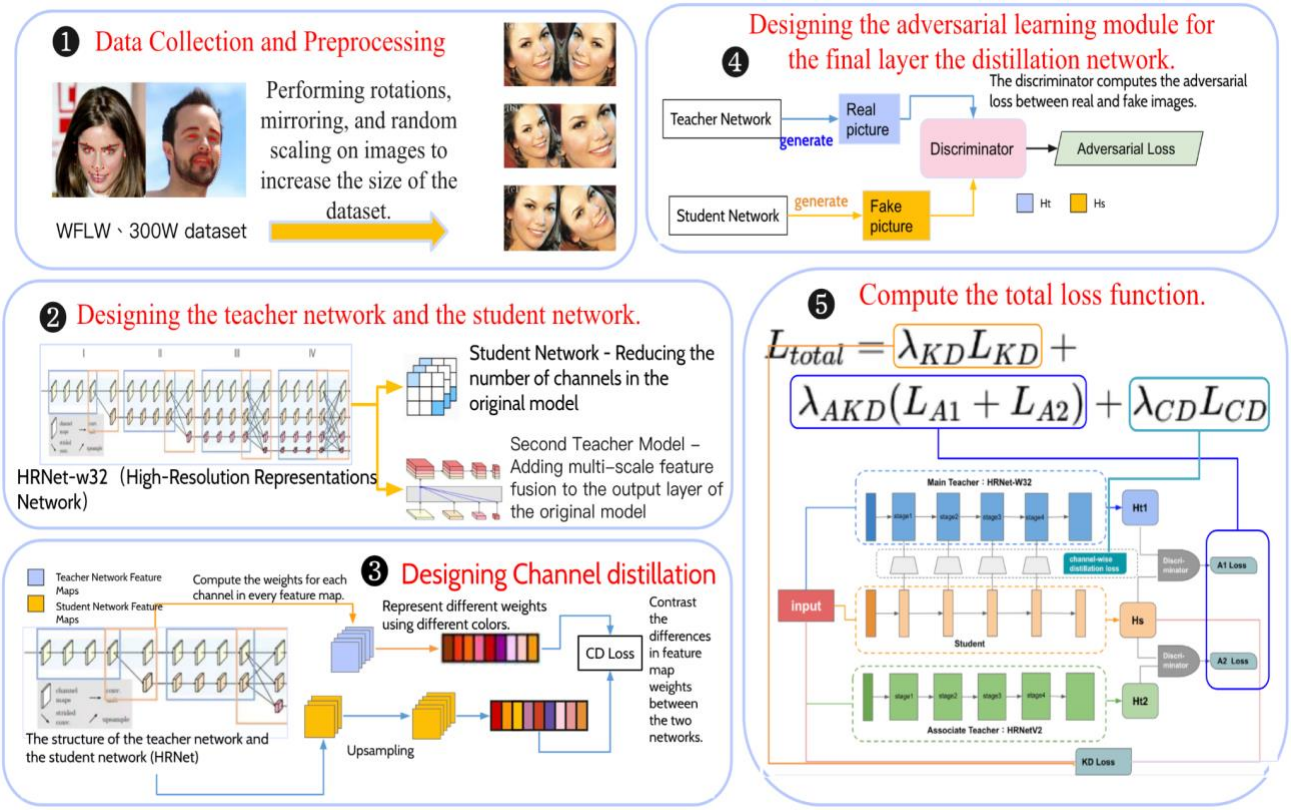

Figure 2 Schematic diagram of the research method

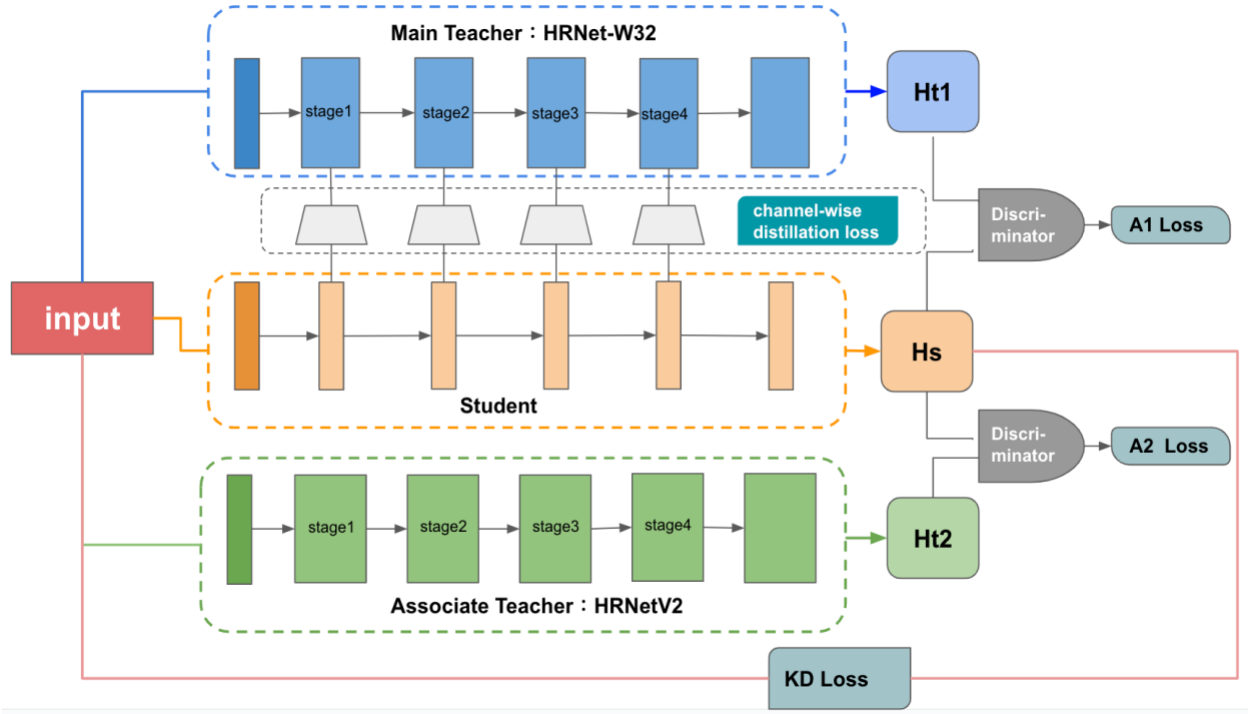

Figure 3 The knowledge distillation structure of this research

5. Expected Results

The future research direction aims to explore a distillation method based on a multi-teacher n etwork, combining channel distillation and adversarial learning techniques, for lightweighting the HR-Net and achieving high-accuracy facial landmark detection algorithms. The expected outcome s include:

1) Implementation of a distillation method based on a multi-teacher network, achieving model par ameter reduction and computational complexity reduction by pruning channels from the original la rge model, thereby improving the operational efficiency and inference speed.

2) Further enhancement of the model's generalization and resistance to interference through the application of adversarial learning techniques.

3) Improved accuracy and robustness of the facial landmark detection model through attention an d integration of channel information.

4) Evaluation of the proposed method's performance, comparison with existing lightweight facial l andmark detection models, and comparison with future research directions in facial landmark det ection models to demonstrate the superiority of the proposed method.

6. References

[1] Newell, A., Yang, K., & Deng, J. Stacked hourglass networks for human pose estimation. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14. Springer International Publishing, 2016. pp. 483-499.

[2] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 770-778.

[3] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2020, 43.10: 3349-3364.

[4] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 4510-4520.

[5] Zhang, X., Zhou, X., Lin, M., & Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 6848-6856.

[6] Dong, X., Yan, Y., Ouyang, W., & Yang, Y. Style aggregated network for facial landmark detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 379-388.

[7] Feng, Z. H., Kittler, J., Awais, M., Huber, P., & Wu, X. J. Wing loss for robust facial landmark localisation with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 2235-2245.

[8] Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., & Zhou, Q. Look at boundary: A boundary-aware face alignment algorithm. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 2129-2138.

[9] Hinton, G., Vinyals, O., & Dean, J. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.

[10] Zhou, Z., Zhuge, C., Guan, X., & Liu, W. Channel distillation: Channel-wise attention for knowledge distillation. arXiv preprint arXiv:2006.01683, 2020.

[11] You, S., Xu, C., Xu, C., & Tao, D. Learning from multiple teacher networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017. pp. 1285-1294.

[12] Goodfellow, I. J., Shlens, J., & Szegedy, C. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572, 2014.

[13] Fard, A. P., & Mahoor, M. H. Facial landmark points detection using knowledge distillation-based neural networks. Computer Vision and Image Understanding, 2022, 215: 103316.

[14] Burgos-Artizzu, X. P., Perona, P., & Dollár, P. Robust face landmark estimation under occlusion. In: Proceedings of the IEEE international conference on computer vision. 2013. pp.1513-1520.

[15] Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In: Proceedings of the IEEE international conference on computer vision workshops. 2013. pp.397-403.