1. Organization of each part:
2. Introduction:
3. Paragraph 1: DNN is good, but large; so need to compress; (2022-1-8)
4. Paragraph 2: small network is hard to train; so need distillation; (2022-1-8)
5. Paragraph 3: our method; (2022-1-8)
6. Hahn, Sangchul, and Heeyoul Choi. "Self-knowledge distillation in natural language processing." arXiv preprint arXiv:1908.01851 (2019). (2022-1-11)
7. Ji, Mingi, et al. "Refine Myself by Teaching Myself: Feature Refinement via Self-Knowledge Distillation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. 文笔太差了; (2022-1-10)
8. Kim, Kyungyul, et al. "Self-knowledge distillation with progressive refinement of targets." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. (2022-1-11)
9. Tung, Frederick, and Greg Mori. "Similarity-preserving knowledge distillation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. (2022-1-10)
10. Zhang, Linfeng, et al. "Be your own teacher: Improve the performance of convolutional neural networks via self-distillation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. (2022-1-10)
11. Zhang, Ying, et al. "Deep mutual learning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. (2022-1-7)
12. Related Work:
13. (1) Knowledge distillation:
14. Knowledge Distillation（知识蒸馏）Review--20篇paper回顾; (2022-1-7)
15. Hou, Yuenan, et al. "Learning lightweight lane detection cnns by self attention distillation." Proceedings of the IEEE/CVF international conference on computer vision. 2019. (2022-1-10)
16. (2) Self attention:
17. Lee, Junhyun, Inyeop Lee, and Jaewoo Kang. "Self-attention graph pooling." International Conference on Machine Learning. PMLR, 2019.
18. Zhang, Han, et al. "Self-attention generative adversarial networks." International conference on machine learning. PMLR, 2019; done; (2022-1-7)
19. Zhao, Hengshuang, Jiaya Jia, and Vladlen Koltun. "Exploring self-attention for image recognition." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. ????? this self-attention is related to convolution, not our purpose; (2022-1-8)
20. <https://lilianweng.github.io/lil-log/2018/06/24/attention-attention.html> (2022-1-8)
21. (3) Ensemble learning:
22. Dong, Xibin, et al. "A survey on ensemble learning." Frontiers of Computer Science 14.2 (2020): 241-258. (2022-1-8)
23. III. Methodology
24. identify the importance of utilizing bottleneck knowledge and introduce the ensemble of intermediate layers’ prediction for improving robustness and effectiveness; introduce our SASD method inspired by self-attention; our approach can be regarded as a kind of ensemble, which obviously will avoid the model overconfidence; (2022-1-12)
25. III.1 self-attention
26. traditional knowledge distillation; temperature hyperparameter is introduced to control the importance of each soft target; (2022-1-12)
27. self-knowledge distillation methods using their own knowledge have been proposed; BYOT; (2022-1-12)
28. Introduce BYOT; (2022-1-12)
29. III.2 our approach
30. directly adds the loss between each intermediate layer feature and the final layer feature without considering the different distillation importance of each bottleneck information, which may cause the loss of valuable information;
31. Experiment:
32. Yun, Sukmin, et al. "Regularizing class-wise predictions via self-knowledge distillation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. (2022-1-7)
33. Conclusion:
34. -