

Semi-Supervised Learning for Glomerular Crescent Classification

Joacy M. da Silva*, Washington L. C. dos Santos[†], Angelo A. Duarte*, Luciano R. de Oliveira[‡],
Michele F. Angelo* and Angelo C. Loula*

*State University of Feira de Santana (UEFS), Bahia, Brazil

Email: joacym Silva@gmail.com, angeloduarte@uefs.br, mfangelo@ecom.uefs.br and angelocl@uefs.br

[†]Gonçalo Moniz Research Center (CpqGM/FIOCRUZ), Bahia, Brazil

Email: wluis@bahia.fiocruz.br

[‡]Federal University of Bahia (UFBA), Bahia, Brazil

Email: lrebouca@ufba.br

Abstract—Glomeruli, the fundamental structures of the kidney responsible for blood filtration, are susceptible to various pathologies, including glomerular crescents. This work investigates the application of semi-supervised deep learning (SSL) for the classification of histopathological images containing these lesions. We evaluate several configurations of a deep learning model for classifying crescent versus normal glomeruli. Our investigation specifically compares the impact of varying the ratio of labeled to unlabeled data and the effect of different input perturbations (e.g., Random Flip + Random Crop) on model performance. Based on accuracy, precision, sensitivity, and F1-score, the model trained with four labeled folds and using Random Flip + Random Crop perturbations emerged as the most effective configuration. In a separate experiment, we expanded the classification task to distinguish crescent from non-crescent glomeruli by adding images of three other lesion types to the dataset. Our results show that a combined training approach using Mean-Teacher and Supervised Contrastive Learning (SCL) did not yield significant performance improvements compared to the model trained solely with the Mean-Teacher method.

I. INTRODUCTION

In nephropathology, glomeruli are key diagnostic indicators for various renal diseases. Consequently, the evaluation of glomerular lesions in histological slides is a critical step in the diagnosis of numerous renal pathologies [1]. One such significant lesion is the glomerular crescent, which is characterized by the proliferation of cells in at least 10% of Bowman’s space. The presence of these crescents in renal biopsies is a hallmark of crescentic glomerulopathies [2].

Several studies have leveraged computational approaches for the analysis of glomerular pathologies. For instance, Uchino et al. [3] trained numerous deep learning models for the classification of seven distinct glomerular lesions, including crescents. Kawazoe et al. [4] developed an automated pipeline for the detection and segmentation of glomeruli and their key histopathological components, including Bowman’s space and crescents, in whole slide images (WSIs) stained with Periodic Acid Schiff (PAS). Similarly, Da Silva et al. [5] evaluated various deep learning models for the classification of crescentic glomeruli, addressing the challenge of class imbalance. Akatsuka et al. [6] also contributed to this area by building a deep learning model for the automatic detection of

glomerular lesions, including crescents, specifically in mouse kidneys. However, of the works mentioned, only that of Li et al. [7] utilized semi-supervised learning (SSL) to address the task of recognizing glomerular lesions, including crescents, particularly in datasets with noisy labels. This highlights a gap in the application of SSL for this specific problem.

While properly trained machine learning models can achieve high accuracy in classification tasks, their performance is contingent on the availability of large, reliably annotated datasets. In digital pathology, however, annotating whole slide images (WSIs) is a time-consuming and labor-intensive process for pathologists [8]. This scarcity of large, well-annotated datasets presents one of the most significant challenges for developing effective AI methods in the field [9].

To address this limitation, Semi-supervised Learning (SSL) has emerged as a promising training strategy [8]–[11]. In this paradigm, the dataset is composed of a labeled dataset (D_l) and an unlabeled dataset (D_u). The labeled data serves to guide the learning process, while the unlabeled data helps the model to better generalize the underlying data distribution [10], [11].

The purpose of this work is to evaluate various configurations for training models to classify images containing glomerular crescents. This is achieved by combining semi-supervised deep learning (SSL), specifically using the Mean-Teacher (MT) method [12], with Transfer Learning (TL) by fine-tuning a pre-trained model. We consider two distinct classification scenarios: a binary task of crescent vs. normal, and a multi-class task of crescent vs. non-crescent (encompassing normal and all other lesion types). The second scenario introduces a significant class imbalance problem. To address this, we evaluate the effectiveness of Suppressed Consistency Loss (SCL) [13].

II. DATASET, TRAINING STRATEGIES AND MODELS

A. Dataset

A dataset of 6507 histological images of human kidney glomeruli from the Gonçalo Moniz Research Center (CpqGM/FIOCRUZ) was used in this study. The dataset comprises five lesion classes: crescents (1,189), normal (1,571),

TABLE I
DISTRIBUTION OF IMAGES BY LESION AND STAIN

Glomerulus	AZAN	H&E	PAMS	PAS	PS	PICRO	Total
Crescent	122	467	157	355	0	88	1189 (18,27%)
Normal	126	869	189	294	0	93	1189 (18,27%)
Hypercellularity	121	623	0	342	5	0	1091 (16,77%)
Sclerosis	178	464	155	298	77	0	1172 (18,01%)
Membranous	0	0	0	0	1484	0	1484 (22,81%)
Total	547 (8,40%)	2423 (37,24%)	501 (7,70%)	1289 (19,81%)	1566 (24,07%)	181 (2,78%)	6507 (100,00%)

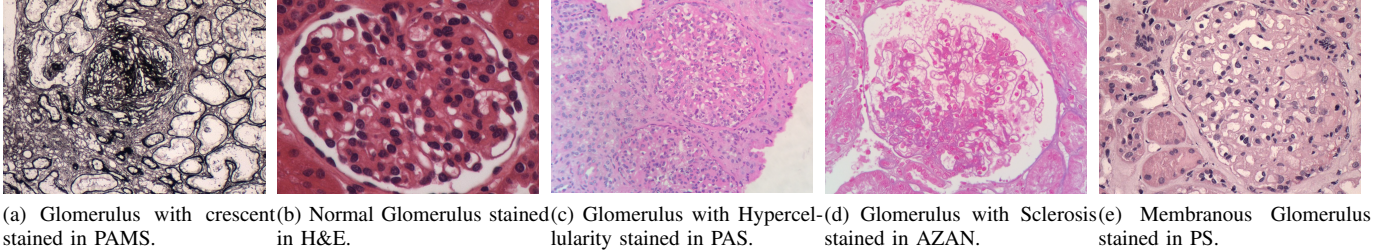


Fig. 1. Examples of glomerulus, with different stains, present in the dataset.

sclerosis (1,172), hypercellularity (1,091), and membranous glomeruli (1,484). Images, acquired with a microscope-mounted digital camera, lack standardized acquisition protocols. Images were stained with various methods, including AZAN, Hematoxylin & Eosin (H&E), Periodic Acid Methenamine Silver (PAMS), Periodic Acid Schiff (PAS), Periodic Schiff (PS), and PICRO. The distribution of images by lesion and stain is shown in Table I, and illustrative examples in Figure 1.

The original images exhibit significant size variability, with dimensions ranging from 205×238 to 3286×4128 pixels. To minimize computational overhead and standardize the input for our experiments, all images were resized to a uniform dimension of 224×224 pixels.

B. Training Strategies and Models

The purpose of this work was to evaluate different training conditions for Deep SSL + TL models in the classification tasks of images of glomerular crescent. Models were trained and evaluated in the classification of crescent vs normal, and crescent vs non-crescent (normal + all other lesions, sclerosis, hypercellularity, and membranous).

Following training strategies outlined by [14] and [15], we defined our experimental conditions by varying the amount of labeled data (10%, 20%, and 40%) and the types of input perturbations. The Mean-Teacher (MT) algorithm [12] was selected for its proven effectiveness in Semi-supervised Learning (SSL). Based on the work of [13], we evaluated combinations of input perturbations, including Random Flip (F), Random Flip + Random Crop (FC), and Random Flip + Random Crop + Gaussian Noise (FCN), to promote model consistency.

We designed models using a pre-trained Wide ResNet-50-2 [16]. Two classification tasks were performed: Crescent vs. Normal, distinguishing between images of crescentic and normal glomeruli; Crescent vs. Non-Crescent, classifying crescent images against all other lesion types (normal, hypercellularity, sclerosis, and membranous). This second task presents a

significant class imbalance, with 1,189 crescent images versus 5,318 non-crescent images. To address this imbalance during Mean-Teacher (MT) [12] training, we applied Suppressed Consistency Loss (SCL) [13].

Pytorch framework was used, with networks trained for 200 epochs with SGD optimizer, the batch size of 64, learning rate of 0.1, weight decay 10^{-4} , momentum of 0.9, exponential moving average 0.95, and consistency coefficient of 8. The experiments were performed in a machine configured with 32 GB of RAM and an NVIDIA GeForce RTX 3080 GPU, with CUDA Version 12.3.

Models' performance measures in the experiments used 10-fold cross-validation (10CV). Images of each glomeruli class were stratified by staining and distributed among folds. Three fold division configurations were considered, according to the sizes of D_l and D_u . The experiments were performed with D_l of sizes 1, 2, and 4 folds, and consequently, D_u of sizes 7, 6, and 4 folds. In each configuration, one fold was used as the test set and one fold as the validation set. A baseline model, trained in a fully supervised manner, was also considered. The Wilcoxon test (95% CI) was compared the performance of the models in the glomeruli classification task. For the crescent vs. normal classification, the measures was Accuracy (acc), Precision (prec), Sensitivity (sens) and F1-score (f1). For the crescent vs. non-crescent classification, the measures Specificity (spec), Geometric Mean (gm) and Matthews Correlation Coefficient (mcc) was also considered.

III. RESULTS AND DISCUSSION

A. Crescent vs Normal Classification

1) *Evaluating the amount of D_l e D_u* : Three fold division configurations were considered, according to the number of folds that would represent D_l and D_u . The experiments were performed with D_l of sizes 1, 2, and 4 folds, and consequently, D_u of sizes 7, 6, and 4 folds. The input perturbation was the Random Flip. Table II shows the results. The model name follows the format: {TRAINING} / PERTURBATIONS / {FOLDS_ D_l } / {FOLDS_ D_u }.

TABLE II
CLASSIFICATION RESULTS EVALUATING THE AMOUNT OF D_l AND D_u .

Training	Accuracy	Precision	Sensibility	F1-score
Sup/F/8/0	0.983 $\pm$ 0.010	0.985 $\pm$ 0.012	0.975 $\pm$ 0.019	0.980 $\pm$ 0.011
MT/F/1/7	0.944 $\pm$ 0.023	0.935 $\pm$ 0.051	0.940 $\pm$ 0.029	0.936 $\pm$ 0.024
MT/F/2/6	0.971 $\pm$ 0.018	0.971 $\pm$ 0.033	0.962 $\pm$ 0.024	0.966 $\pm$ 0.021
MT/F/4/4	0.981 $\pm$ 0.010	0.986 $\pm$ 0.014	0.970 $\pm$ 0.026	0.977 $\pm$ 0.012

Regarding acc, the {MT/F/4/4} model significantly outperformed the {MT/F/1/7} ($p < 0.001$) and {MT/F/2/6} ($p = 0.043$) models. The same can be said for prec ($p < 0.001$ and $p = 0.010$, respectively). Compared to the {MT/F/1/7} model, the {MT/F/4/4} model also proved to be superior in the other measures, sens ($p = 0.014$) and f1 ($p < 0.001$). The same cannot be said for sens ($p = 0.157$) and f1 ($p = 0.053$) compared to the {MT/F/2/6} model.

2) *Evaluating the input perturbations:* The experiments were performed with D_l of sizes 1, 2, and 4 folds, and consequently, D_u of sizes 7, 6, and 4 folds. Table III presents the results. The model name follows the format: {TRAINING} / PERTURBATIONS / {FOLDS_ D_l } / {FOLDS_ D_u }.

TABLE III
CLASSIFICATION RESULTS EVALUATING THE INPUT PERTURBATIONS.

Training	Accuracy	Precision	Sensibility	F1-score
MT/F/1/7	0.944 $\pm$ 0.023	0.935 $\pm$ 0.051	0.940 $\pm$ 0.029	0.936 $\pm$ 0.024
MT/FC/1/7	0.960 $\pm$ 0.020	0.954 $\pm$ 0.037	0.955 $\pm$ 0.025	0.954 $\pm$ 0.022
MT/FCN/1/7	0.944 $\pm$ 0.024	0.965 $\pm$ 0.035	0.903 $\pm$ 0.035	0.932 $\pm$ 0.030
MT/F/2/6	0.971 $\pm$ 0.018	0.971 $\pm$ 0.033	0.962 $\pm$ 0.024	0.966 $\pm$ 0.021
MT/FC/2/6	0.974 $\pm$ 0.014	0.975 $\pm$ 0.016	0.964 $\pm$ 0.022	0.969 $\pm$ 0.016
MT/FCN/2/6	0.960 $\pm$ 0.015	0.988 $\pm$ 0.007	0.919 $\pm$ 0.037	0.952 $\pm$ 0.019
MT/F/4/4	0.981 $\pm$ 0.010	0.986 $\pm$ 0.014	0.970 $\pm$ 0.026	0.977 $\pm$ 0.012
MT/FC/4/4	0.982 $\pm$ 0.008	0.987 $\pm$ 0.013	0.971 $\pm$ 0.019	0.979 $\pm$ 0.010
MT/FCN/4/4	0.966 $\pm$ 0.014	0.985 $\pm$ 0.014	0.934 $\pm$ 0.034	0.959 $\pm$ 0.018

Considering only models trained with 1 fold composing the D_l , through a two-tailed test it can be verified that there was a tie in the acc ($p = 0.953$) and f1 ($p = 0.859$) measures between the {MT/F/1/7} model and the {MT/FCN/1/7} model. The {MT/F/1/7} model presented a significantly superior result in sens ($p = 0.033$). The {MT/FC/1/7} model presented significantly superior results in the measures of acc ($p = 0.010$), sens ($p < 0.001$), and f1 ($p = 0.005$) compared to the {MT/FCN/1/7} model, and in the measures of acc ($p = 0.033$) and f1 ($p = 0.033$) compared to the {MT/F/1/7} model. Considering prec ($p = 0.221$) and sens ($p = 0.157$), compared to the {MT/F/1/7} model, it cannot be stated that the results obtained by the {MT/FC/1/7} model were significantly superior. Regarding prec, it cannot be stated that the result obtained by the {MT/FCN/1/7} model was superior to that of the other two models.

Regarding the models trained with 2 folds composing the D_l , the {MT/FC/2/6} model presented significantly better performance than the {MT/FCN/2/6} model in the measures of acc ($p = 0.005$), sens ($p = 0.004$) and f1 ($p = 0.005$). The {MT/F/2/6} model also presented significantly better performance compared to the {MT/FCN/2/6} model in the

measures of acc ($p = 0.025$), sens ($p = 0.007$) and f1 ($p = 0.024$). However, when the measure in question was prec, the {MT/FCN/2/6} model presented significantly better results than the {MT/F/2/6} and {MT/FC/2/6} models. The two-sided test was applied to evaluate whether the measurements obtained with the {MT/F/2/6} and {MT/FC/2/6} models presented differences. The results showed that there was a tie between the two models, in the four measurements evaluated.

Considering the context of 4 labeled folds, the {MT/FC/4/4} model significantly outperformed the {MT/FCN/4/4} model in the measures of acc ($p < 0.001$), sens ($p < 0.001$), and f1 ($p < 0.001$). The two-sided test was applied to evaluate whether the measures obtained with the {MT/FC/4/4} and {MT/F/4/4} models presented differences. The results showed that there was a tie between the two models in the four measures evaluated. The {MT/F/4/4} model also significantly outperformed the {MT/FCN/4/4} model in the measures of acc ($p < 0.001$), sens ($p = 0.004$), and f1 ($p < 0.001$). However, when the measurement in question was prec, the test result indicated a tie between the three models.

As in the work by [15], when including Gaussian Noise in the input perturbations, the model showed a considerable drop in performance in most measures. In the context of this work, it was found that only the prec measure did not significantly lower the value compared to the other models. The experiment confirms the influence of perturbations on model performance and may indicate that Random Flip + Random Crop can be adopted as the perturbation applied to the inputs in future experiments, since it, together with Random Flip, obtained the best results in the experiments for a D_l composed of 2 and 4 folds, and superior acc and f1 measures for a D_l composed of only 1 fold.

B. Crescent vs Non-crescent Classification

For the crescent vs. non-crescent classification task, the classification must distinguish between crescent and non-crescent images, which include normal, glomerulus with hypercellularity, sclerosis, and membranous. Considering the imbalance of the dataset, SCL [13] was integrated into semi-supervised training with MT [17] in an attempt to minimize the problem.

The experiments proposed in sections III-A1 and III-A2 were carried out with the purpose of finding the configuration, in relation to amount of D_l and D_u , and the input perturbations used as data augmentation. For this experiment, the configuration adopted in relation to the amount of D_l and D_u was that of 4 folds for each set. The input perturbation adopted was the combination of Random Flip and Random Crop. Table IV presents the results of the classifications performed by the models. The model name follows the format: {TRAINING} / PERTURBATIONS / {FOLDS_ D_l } / {FOLDS_ D_u }.

The results obtained in this classification reaffirm the impact of the class imbalance, and the tendency to classify the majority class. After applying the statistical test, it was not possible to confirm that the {MT+SCL=0.2/FC/4/4} model presented significantly superior results compared to the {MT/FC/4/4} model in any of the measures.

TABLE IV
RESULTS OF CRESCENT VS NON-CRESCENT CLASSIFICATION.

Training	Accuracy	Precision	Sensitivity	F1-score	Specificity	Geometric Mean	MCC
Sup/FC/8/0	0.930 ± 0.010	0.831 ± 0.031	0.773 ± 0.045	0.800 ± 0.030	0.965 ± 0.008	0.863 ± 0.025	0.759 ± 0.034
MT/FC/4/4	0.924 ± 0.015	0.846 ± 0.035	0.713 ± 0.083	0.771 ± 0.055	0.971 ± 0.008	0.831 ± 0.049	0.732 ± 0.058
MT+SCL=0.2/FC/4/4	0.929 ± 0.010	0.843 ± 0.040	0.754 ± 0.024	0.796 ± 0.027	0.968 ± 0.009	0.854 ± 0.016	0.755 ± 0.034

IV. CONCLUSION

We evaluated various Deep Semi-supervised Learning (SSL) models for two distinct glomerular crescent classification tasks. Our training strategies varied the amount of labeled data and the types of input perturbations. In the crescent vs. normal task, increasing the amount of labeled data led to improved performance. Models trained with Random Flip and Random Crop perturbations consistently achieved the best results. For the crescent vs. non-crescent task, the models demonstrated a bias towards classifying the majority non-crescent class. Although the application of Suppressed Consistency Loss (SCL) showed an apparent improvement in metrics such as sensitivity, F1-score, geometric mean, and MCC, this enhancement was not statistically significant compared to the baseline Mean-Teacher (MT) model without SCL.

Future work will involve several key comparisons. First, we plan to compare our Semi-supervised Learning (SSL) models with those trained in a fully supervised manner. Second, we will benchmark the Mean-Teacher (MT) algorithm against other SSL methods, such as the hybrid approach FixMatch [18]. Finally, we intend to explore alternative approaches for addressing class imbalance, including algorithms like CREST [19] and DASO [20].

REFERENCES

- [1] P. Chagas, L. Souza, I. Araújo, N. Aldeman, A. Duarte, M. Angelo, W. L. dos Santos, and L. Oliveira, "Classification of glomerular hypercellularity using convolutional features and support vector machine," *Artificial Intelligence in Medicine*, vol. 103, p. 101808, Mar. 2020. [Online]. Available: <https://doi.org/10.1016/j.artmed.2020.101808>
- [2] L. Anguiano, R. Kain, and H.-J. Anders, "The glomerular crescent: triggers, evolution, resolution, and implications for therapy," *Current Opinion in Nephrology and Hypertension*, vol. 29, no. 3, pp. 302–309, May 2020. [Online]. Available: <https://doi.org/10.1097/mnh.0000000000000596>
- [3] E. Uchino, K. Suzuki, N. Sato, R. Kojima, Y. Tamada, S. Hiragi, H. Yokoi, N. Yugami, S. Minamiguchi, H. Haga, M. Yanagita, and Y. Okuno, "Classification of glomerular pathological findings using deep learning and nephrologist-AI collective intelligence approach," *International Journal of Medical Informatics*, vol. 141, p. 104231, Sep. 2020. [Online]. Available: <https://doi.org/10.1016/j.ijmedinf.2020.104231>
- [4] Y. Kawazoe, K. Shimamoto, R. Yamaguchi, I. Nakamura, K. Yoneda, E. Shinohara, Y. Shintani-Domoto, T. Ushiku, T. Tsukamoto, and K. Ohe, "Computational pipeline for glomerular segmentation and association of the quantified regions with prognosis of kidney function in IgA nephropathy," *Diagnostics*, vol. 12, no. 12, p. 2955, Nov. 2022. [Online]. Available: <https://doi.org/10.3390/diagnostics12122955>
- [5] J. M. da Silva, M. F. Angelo, W. L. C. dos Santos, and A. C. Loula, "Aprendizado profundo na classificação de lesões crescentes glomerulares: modelos e condições," in *CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 34. (SIBGRAPI)*. Gramado (Virtual), Brasil: Proceedings... Los Alamitos: IEEE Computer Society, 2021, 2021. [Online]. Available: <http://urlib.net/ibi/8JMKD3MGPEW34M/45F7NEB>
- [6] A. Akatsuka, Y. Horai, and A. Akatsuka, "Automated recognition of glomerular lesions in the kidneys of mice by using deep learning," *Journal of Pathology Informatics*, vol. 13, p. 100129, 2022. [Online]. Available: <https://doi.org/10.1016/j.jpi.2022.100129>

- [7] J. Li, Q. He, Y. Liu, Y. Wang, T. Guan, J. Ye, Y. He, and Z. Wang, "Glomerular lesion recognition based on pathology images with annotation noise via noisy label learning," *IEEE Access*, vol. 11, pp. 41 325–41 336, 2023. [Online]. Available: <https://doi.org/10.1109/access.2023.3269792>
- [8] J. U. Becker, D. Mayerich, M. Padmanabhan, J. Barratt, A. Ernst, P. Boor, P. A. Cicalese, C. Mohan, H. V. Nguyen, and B. Roysam, "Artificial intelligence and machine learning in nephropathology," *Kidney International*, vol. 98, no. 1, pp. 65–75, Jul. 2020. [Online]. Available: <https://doi.org/10.1016/j.kint.2020.02.027>
- [9] P.-C. Chung, W.-J. Yang, T.-H. Wu, C.-R. Huang, and Y.-Y. Hsu, "Emerging research directions of deep learning for pathology image analysis," in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, Oct. 2022. [Online]. Available: <https://doi.org/10.1109/biomas54905.2022.9948651>
- [10] Q. Gui, H. Zhou, N. Guo, and B. Niu, "A survey of class-imbalanced semi-supervised learning," *Machine Learning*, vol. 113, no. 8, p. 5057–5086, May 2023. [Online]. Available: <http://dx.doi.org/10.1007/s10994-023-06344-7>
- [11] W. D. G. de Oliveira and L. Berton, "A systematic review for class-imbalance in semi-supervised learning," *Artificial Intelligence Review*, vol. 56, no. S2, p. 2349–2382, Sep. 2023. [Online]. Available: <http://dx.doi.org/10.1007/s10462-023-10579-0>
- [12] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/68053af2923e00204c3ca7c6a3150cf7-Paper.pdf>
- [13] M. Hyun, J. Jeong, and N. Kwak, "Class-imbalanced semi-supervised learning," 2020. [Online]. Available: <https://arxiv.org/abs/2002.06815>
- [14] A. Oliver, A. Odena, C. A. Raffel, E. D. Cubuk, and I. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018. [Online]. Available: <https://proceedings.neurips.cc/paper/2018/file/c1fea270c48e8079d8ddf7d06d26ab52-Paper.pdf>
- [15] H.-Y. Zhou, A. Oliver, J. Wu, and Y. Zheng, "When semi-supervised learning meets transfer learning: Training strategies, models and datasets," 2018. [Online]. Available: <https://arxiv.org/abs/1812.05313>
- [16] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016. [Online]. Available: <https://arxiv.org/abs/1605.07146>
- [17] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol. 30, 2017.
- [18] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 596–608. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/file/06964dce9adbb1c5cb5d6e3d9838f733-Paper.pdf>
- [19] C. Wei, K. Sohn, C. Mellina, A. Yuille, and F. Yang, "Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 10 857–10 866.
- [20] Y. Oh, D.-J. Kim, and I. S. Kweon, "Daso: Distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 9786–9796.