NATIVE MACHINE LEARNING FOR NOISY MICROSCOPIC DATA PROCESSING

Danil Afonchikov, Irina Makovik, Elena Kornaeva
Orel State University
danil.afonchikov@bk.ru, irina.makovik@gmail.com, smkornaeva@gmail.com

Alexey Kornaev
Innopolis University
avkornaev@gmail.com

ABSTRACT

Cells count become a challenging problem when the cells move in a continuous stream, and their boundaries are difficult for visual detection. To resolve this problem we modified the training and decision making processes using curriculum learning and multi-view predictions techniques, respectively.

1 INTRODUCTION

There are two main approaches to low-quality images processing. The first approach deals with image restoration (Schermelleh et al., 2019; Lehtinen et al., 2018; Xu et al., 2020). This research is related to the second approach which deals with robustness of machine learning (Matiisen et al., 2019; Radosavovic et al., 2018; Li et al., 2021; Bachman et al., 2019; Wenzel et al., 2020). Particulary, the research deals with *curriculum learning* (Wang et al., 2021) which is learning from simple tasks to complex ones, and *multi-view* post-processing which allows to make the final prediction based on multiple preliminary predictions. The main application of the research is microscopic image and video processing. One of the ways to study blood cells inside a human body is capillaroscopy (Ingegnoli et al., 2018). Data on blood micro-circulations obtained using non-invasive methods of capillaroscopy normally have relatively low contrast since the observation takes place through soft tissues (Abdou et al., 2022; Dremin et al., 2019). Rare exceptions are the methods that require the use of the most sophisticated equipment and measurement procedures (McKay et al., 2020). Some diseases are accompanied by a significant change in number of blood cells (Deng et al., 2023). An express *in vivo* blood count test that estimates at least low/high level of blood cells is required in hematology.

2 METHODOLOGY

The paper deals with an original synthetic video dataset of moving red and white blood cells (RBCs, WBCs, respectively). The proposed dataset is an attempt to connect the existing *in vitro* complete blood count tests (see appendix A, fig. 4a) (Shenggan, 2017) and the prospective *in vivo* tests using capillaroscopy data (see appendix A, fig. 1e) (Herrick et al., 2021; Gurov et al., 2018). Some examples of a nailfold capillaroscopy and synthetic dataset videos are given in the supplementary materials. The proposed synthetic dataset contains 1150 videos with 100 frames in each video of $3 \times 128 \times 128$ frames. See appendix A for detailed information on the dataset. The paper uses competence-based *curriculum learning* (CL) to reduce training time and increase accuracy of the models (Platanios et al., 2019). The main idea of the method is to calculate difficulty of each sample and use the samples with difficulty score less or equal to the competence at a given training epoch (see Appendix B, Equation (2)). The proposed *multi-view* (MV) method refers to the idea of ensembling and using a set of slightly different networks that make a set of predictions to estimate the uncertainty of the predictions (Ashukha et al., 2020). But in work, a set of slightly different (augmented) variants of an input is used to make a set of preliminary predictions. It is supposed

Table 1: Comparison of the baselines and the proposed models based on curriculum learning (CL) and multi-view (MV) predictions. The synthetic video dataset of moving blood cells was used.

MODEL	**#PARAM.** (train. par.)	**#EP.**	**INPUT**	**ACCURACY, %** Baseline	MV / CL
			White blood cells (WBC)		
EffNet pr.	3.8 M (all)	400	[9,3,64,64]	$\mathbf{83.56 \pm 3.05}$	$\mathbf{90.26 \pm 0.89}/-$
9-layer CNN	4.4 M (all)	800	[3,32,32]	81.57 ± 3.32	$89.39 \pm 2.58/-$
ResNet18 pr.	11.2 M (all)	400	[3,256,256]	82.61 ± 2.53	$87.83 \pm 2.70/-$
ViT pr.	303 M (2 K)	400	[3,224,224]	76.31 ± 2.76	$85.00 \pm 2.15/-$
			Red blood cells (RBC)		
EffNet pr.	3.8 M (all)	400	[3,64,64]	62.6 ± 0.53	$-/62 \pm 0.51$
EffNet pr.	3.8 M (all)	200	[9,3,64,64]	64.6 ± 0.52	$-/65.6 \pm 0.52$

that the final prediction is the most frequent preliminary prediction, or prediction with the highest cumulative weight (see Appendix B, Equation (4), Equation (5), respectively).

3 Results and discussion

Two series of experiments were performed (see Appendix A, Appendix C). The *first series* demonstrated that the proposed multi-view approach increases the accuracy significantly, which is partially the result of a relatively small crop area of the augmented frames (see Table 1). The number of multi-views and the parameters of data augmentation, e.g. random crop scale, are the additional hyperparameters of the multi-view approach. A set of preliminary tests demonstrated that the more the number of multi-views, the more the accuracy of the model. But, the dependence on the crop scale is not obvious and should be studied in further work. The best accuracy of $90.26 \pm 0.89\%$ was obtained with EffNet X3D, the network with relatively small number of trainable parameters, and the only network that processed a short 9 frame videos instead of single frames (see Table 1). ResNet18 had the highest number of trainable parameters, and the model demonstrated a tendency to overfitting which is the result of using a relatively small dataset (see Appendix C, Figure 4). The *the second series* dealt with a more difficult task, since the variance of RBCs distribution in the dataset is small compared to its mean value (see Appendix A).Accuracy of $62 \pm 0.51\%$ with single-frame version of the dataset and accuracy of $65.6 \pm 0.52\%$ with 100-frames version of the dataset using 9 randomly selected consecutive frames were obtained. The effect of curriculum learning is not obvious, since the method increased the accuracy of the model for the video processing by about 1%, but decreased the accuracy of the single-frame model. In both series, the use of videos of moving blood cells allowed to increase the accuracy of the models by about 2-4%. The multi-veiw predictions demonstrated good results in all the tests. The method follows the principle of reductionism when the properties of a part correspond to the properties of the whole object. MV scans an image with relatively small random crops and accumulates the information on predictions, which is probably close to the idea of convolutional networks, but implemented at the post processing stage. The method was additionally tested in processing of noisy data in a public dataset (see Appendix D).

4 Conclusion

Some of the real-world cells datasets are not applicable for cells detection and labeling, but the task of qualitative or quantitative estimation of the cells number remains actual. The obtained results demonstrated that the number of relatively big and low concentrated white blood cells can be estimated as high or low with the accuracy of about 90%. But the number of relatively small and high concentrated red blood cells cannot be estimated accurately, the obtained result in accuracy of about 66% is close to random predictions. The methods of curriculum learning and multi-view predictions are easy in implementation and promising in application to processing of relatively small and low quality datasets. The further research should be related to processing of the real-world data obtained with *in vivo* methods of blood cells visualization. Data collection for a case of well-separated blood count test results and solution of the binary classification problems are of interests.

ACKNOWLEDGEMENT

This research has been financially supported by The Analytical Center for the Government of the Russian Federation (Agreement No. 70-2021-00143 dd. 01.11.2021, IGK 000000D730324P540002). Authors also express gratitude to R&D Center of Biomedical Photonics at Orel State University for sharing their knowledge in the field of microscopic imaging.

URM STATEMENT

We acknowledge that all the authors meet the URM criteria of ICLR 2024 Tiny Papers Track.

REFERENCES

Maged Abdalla Helmy Mohamed Abdou, Trung Tuyen Truong, Eric Jul, and Paulo Ferreira. Deep learning and computer vision techniques for microcirculation analysis: A review, 2022.

Arsenii Ashukha, Alexander Lyzhov, Dmitry Molchanov, and Dmitry Vetrov. Pitfalls of in-domain uncertainty estimation and ensembling in deep learning. *arXiv preprint arXiv:2002.06470*, 2020.

Philip Bachman, R Devon Hjelm, and William Buchwalter. Learning representations by maximizing mutual information across views. *Advances in neural information processing systems*, 32, 2019.

Shuling Deng, Mingyi Lv, and Xiaofeng Long. Correlation between pulmonary embolism and blood cell count changes. *Cellular and Molecular Biology*, 69:63–66, 2023. ISSN 1165158X. doi: 10.14715/cmb/2023.69.10.8.

Viktor Dremin, Igor Kozlov, Mikhail Volkov, Nikita Margaryants, Andrey Potemkin, Evgeny Zherebtsov, Andrey Dunaev, and Igor Gurov. Dynamic evaluation of blood flow microcirculation by combined use of the laser doppler flowmetry and high-speed videocapillaroscopy methods. *Journal of biophotonics*, 12(6):e201800317, 2019.

Christoph Feichtenhofer. X3d: Expanding architectures for efficient video recognition, 2020.

Igor Gurov, Mikhail Volkov, Nikita Margaryants, Aleksei Pimenov, and Andrey Potemkin. High-speed video capillaroscopy method for imaging and evaluation of moving red blood cells. *Optics and Lasers in Engineering*, 104:244–251, 2018.

Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels, 2018.

Ariane L Herrick, Michael Berks, and Chris J Taylor. Quantitative nailfold capillaroscopy—update and possible next steps. *Rheumatology*, 60(5):2054–2065, 2021.

Francesca Ingegnoli, Vanessa Smith, Alberto Sulli, and Maurizio Cutolo. Capillaroscopy in routine diagnostics: potentials and limitations. *Current Rheumatology Reviews*, 14(1):5–11, 2018.

Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009. URL `https://api.semanticscholar.org/CorpusID:18268744`.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

Junnan Li, Caiming Xiong, and Steven C.H. Hoi. Learning from noisy data with robust representation learning. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9465–9474, 2021. doi: 10.1109/ICCV48922.2021.00935.

Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey. Normalized loss functions for deep learning with noisy labels, 2020.

Tambet Matiisen, Avital Oliver, Taco Cohen, and John Schulman. Teacher–student curriculum learning. *IEEE transactions on neural networks and learning systems*, 31(9):3732–3740, 2019.

Gregory N McKay, Nela Mohan, and Nicholas J Durr. Imaging human blood cells in vivo with oblique back-illumination capillaroscopy. *Biomedical optics express*, 11(5):2373–2382, 2020.

Tim Pearce, Alexandra Brintrup, and Jun Zhu. Understanding softmax confidence and uncertainty. *CoRR*, abs/2106.04972, 2021. URL `https://arxiv.org/abs/2106.04972`.

Emmanouil Antonios Platanios, Otilia Stretcu, Graham Neubig, Barnabas Poczos, and Tom M. Mitchell. Competence-based curriculum learning for neural machine translation, 2019.

Ilija Radosavovic, Piotr Dollár, Ross Girshick, Georgia Gkioxari, and Kaiming He. Data distillation: Towards omni-supervised learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4119–4128, 2018.

Lothar Schermelleh, Alexia Ferrand, Thomas Huser, Christian Eggeling, Markus Sauer, Oliver Biehlmaier, and Gregor PC Drummen. Super-resolution microscopy demystified. *Nature cell biology*, 21(1):72–84, 2019.

Shenggan. Bccd dataset, 2017. URL `https://github.com/Shenggan/BCCD_Dataset`.

Xin Wang, Yudong Chen, and Wenwu Zhu. A survey on curriculum learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4555–4576, 2021.

Jiaheng Wei, Hangyu Liu, Tongliang Liu, Gang Niu, Masashi Sugiyama, and Yang Liu. To smooth or not? when label smoothing meets noisy labels. *arXiv preprint arXiv:2106.04149*, 2021.

Florian Wenzel, Jasper Snoek, Dustin Tran, and Rodolphe Jenatton. Hyperparameter ensembles for robustness and uncertainty quantification. *Advances in Neural Information Processing Systems*, 33:6514–6527, 2020.

Xiaobo Xia, Tongliang Liu, Bo Han, Mingming Gong, Jun Yu, Gang Niu, and Masashi Sugiyama. Sample selection with uncertainty of losses for learning with noisy labels. *CoRR*, abs/2106.00445, 2021. URL `https://arxiv.org/abs/2106.00445`.

Jun Xu, Yuan Huang, Ming-Ming Cheng, Li Liu, Fan Zhu, Zhou Xu, and Ling Shao. Noisy-as-clean: Learning self-supervised denoising from corrupted image. *IEEE Transactions on Image Processing*, 29:9316–9329, 2020.

A Data collection

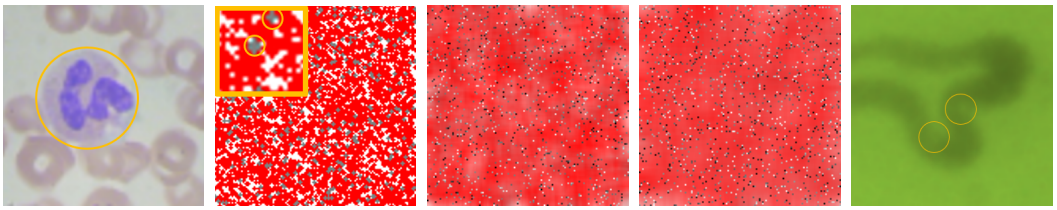
Modern complete blood count (CBC) datasets are based on *in vitro* tests results. The datasets contain high quality images of blood cells. So, object detection methods can be applied to process them. Data obtained with *in vivo* videocapillaroscopy contain images distorted with blur and noise. So object detection methods are not normally applicable. In addition, such distortions worsen the results of AI model predictions, making said models inadequate.

It is supposed that implementation of native machine learning methods including curriculum learning and multi-view predictions in practice of *in vivo* CBC tests can make predictions more reliable. In order to test that, the dataset of blood cells with increasing difficulty factor is needed. A synthetic video dataset of moving blood cells was generated to approach the solution of a problem of *in vivo* CBC tests.

The dataset includes clear, blurred, noisy, and both blurred and noisy samples in proportion $0.130 \times 0.304 \times 0.217 \times 0.349$, respectively. The number RBCs and WBCs have normal distribution in the dataset with mean values of 5000 and 202 and standard deviations of 97.9 and 95.5 respectively. 'High' RBCs or 'high' WBCs levels are used as class '1' in labels. It is supposed that the 'high' level corresponds to the number of cells higher than the mean value. The dataset contains 576 (0.5) videos with 'high' WBCs and 593 (0.512) videos with 'high' RBC levels.

To simulate Brownian motion of blood cells, the Wiener process equation is applied:

$$W\left(\frac{i}{n}\right) = W\left(\frac{i-1}{n}\right) + \frac{Y_i}{\sqrt{n}}. \tag{1}$$



(a) in vitro (b) synthetic, $b = 0$ (c) synthetic, $b = 5$ (d) synthetic, $b = 10$ (e) in vivo

Figure 1: Blood cells microscopy: moving along the path from *in vitro* (a) to *in vivo* (e) blood count tests using the synthetic blood count dataset with different values of blurring b in the samples (b)-(d). Some of the white blood cells situated among the red blood cells are marked with circles (b), (e).

where n is the number of video frames, i is the index of current frame, $Y_i = \pm 1$, determined randomly.

Equation (1) was applied to each coordinate and each cell to determine the moving cell's position at each video frame. Then the box blur was applied and the noise was drawn (see Figure 1).

B Curriculum learning and multi-view predictions intuition

Curriculum learning

Curriculum learning is a machine learning strategy when the model is trained on a sequence of samples and the difficulty of each sample gradually increases. The idea behind this approach is to mimic the way humans learn, by starting with simple concepts and gradually building up to more complex ones. This can help the model to learn more effectively and efficiently, as it is able to build upon its existing knowledge as it progresses through the curriculum.

Competence-based curriculum learning is a specific type of curriculum learning. The competence of the model is evaluated, and in the current learning epoch the samples with difficulty (d) less or equal to competence score ($c(t)$) are used (Platanios et al., 2019). The difficulty estimation (d) and the competence equation ($c(t)$) have the following formalization:

$$d = \alpha b + \beta l, \quad c(t) = \min\left(1, \sqrt[p]{t\frac{1 - c_0^p}{T} + c_0^p}\right), \tag{2}$$

where b is the blur radius for box blurring, $(0 \leq b \leq 10)$, l is the distance from the given target value to the mean value (exact number of cells), α and β are the hyperparameters $(0 \leq \alpha, \beta \leq 1, \alpha + \beta = 1)$, c_0 is the initial competence value, T is the total duration of curriculum learning phase, t is current epoch, $p \geq 1$.

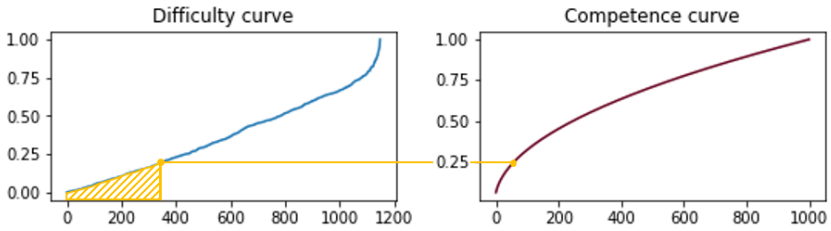


Figure 2: Curriculum learning intuition. The difficulty and competence curves represents Equation (2) when $\alpha = \beta = 0.5$, $c_0 = 0.05$, $T = 1000$, and $p = 2$.

Multi-view

Microscopic images of cells normally include a number of cells and a small part of an image contains approximately the same information on cells condition and concentration as the whole image. So, data augmentation (image cropping, first of all) expands the dataset during the training and makes

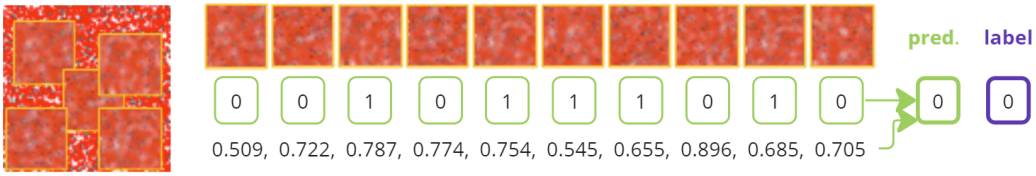


Figure 3: Multi-view intuition. The network recieves a set of fragments of the sample and makes a set of predictions. The final prediction corresponds to the mode value (MVM, see Equation (4)), or to the bin count weighted with confidences (MVWCo-S, see Equation (5)).

a small dataset enough for deep learning models. But also the augmentation allows to average test predictions and estimate their uncertainty using techniques of multi-view predictions.

The goal of multi-view predictions is to use a sample with augmentations multiple times as follows. The inference model receives m augmented copies of each input and calculates m pairs of predicted classes and *confidences* of the predictions $\{y^{(i,j)\,\prime}, c^{(i,j)}\}$ for each input, respectively:

$$y^{(i,j)\,\prime} = \arg\max_k \left[h_k^{(i,j)}\right], \quad c^{(i,j)} = \underset{k}{max}(h_k^{(i,j)}), \tag{3}$$

where k is the class number, (i, j) are the dummy indexes that refer to j^{th} augmented copy of ith sample, $c^{(i,j)}$ is the maximum component of a model prediction vector $\mathbf{h}^{(i,j)}$ which is called confidence, $0 < c^{(i,j)} < 1$.

The confidence values $c^{(i,j)}$ can be interpreted as an certainty or uncertainty estimator of the model predictions (Pearce et al., 2021): the more the confidence value, the more the certainty of the prediction.

There are a few ways to calculate resulting predictions. The simplest way is to calculate mode values over the multi-view predictions (MVM):

$$y^{(i)\,\prime} = \underset{j}{mode}\left[y^{(i,j)\,\prime}\right]. \tag{4}$$

The alternative approach uses information on certainty of the predictions and deals with cumulative confidences:

$$z^{(i)}_{k=y^{(i,j)\,\prime}} = \sum_j c^{(i,j)}, \quad y^{(i)\,\prime} = \max_k(z_k^{(i)}), \tag{5}$$

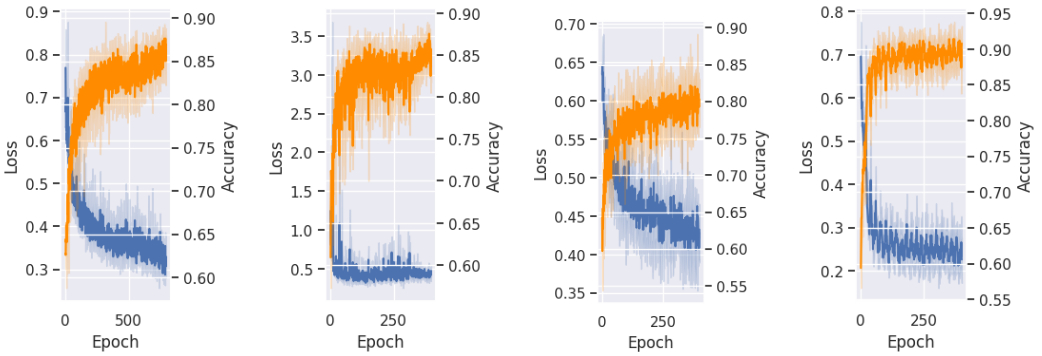
where $\mathbf{z}^{(i)}$ is the additional vector with components $z_k^{(i)}$ which accumulates weights of the predicted classed over the multi-views.

Equation (5) represents bin count of predicted classes $y^{(i,j)\,\prime}$ with weights $c^{(i,j)}$. The method is called soft multi-view predictions weighted with confidences $c^{(i,j)}$ (MVWCo-S).

C Details of experiments

Two series of experiments were performed. In both series the dataset was divided into training, validation and test sets in the proportions of 0.6, 0.2, 0.2, respectively. In both series the random resized crop of $10 - 20\%$ of the frame area was applied for the input data augmentation at the training, validation, and test stages. The Adam optimizer (momentum 0.9) was used with an initial learning rate of 0.001. Five parallel tests with random seeds of $[42, 0, 17, 9, 3]$ were performed, the means and standard deviations were calculated.

The first series of the experiments dealt with various network architectures and with classification of 'high' or 'low' levels of WBCs (higher or lower than the mean value of WBCs). Four network architectures were applied with the following training tactics (see table 1):



(a) 9-layer CNN (b) ResNet18 pretr. (c) ViT pretr. (d) EffNet X3D pretr.

Figure 4: The validation loss (blue lines) and the validation accuracy (orange lines) functions (see Table 1) while the train loss functions demonstrated monotonous decreasing in all the experiments.

- pretrained EffNet X3D with 3.8 million parameters (fine tuned);
- 9-layer CNN (Xia et al., 2021) with 4.4 million parameters (randomly initialized);
- pretrained ResNet18 with 11.2 million parameters (fine tuned);
- pretrained ViT L 16 with 303 million parameters (2 thousand parameters of the 'head' of the model were randomly initialized, the rest of the parameters were frozen).

Ten multi-views (augmented versions) of a test sample were used (see fig. 3). Table 1 and Figure 4 demonstrate the experiments results and the training processes, respectively.

The second series of the experiments dealt with both single-frame version and 100-frames version of the dataset with the classification of 'high' or 'low' level of RBCs. Pretrained Effitient X3D model (Feichtenhofer, 2020) was used. The model was trained with cosine annealing learning rate with starting value of 0.001. Competence-based curriculum learning (Platanios et al., 2019) was used. The model was trained in 400 epochs with single-frame version of the dataset and in 200 epochs with 100-frames version of the dataset using 9 randomly selected consecutive frames (see Table 1).

D Additional experiments on CIFAR-10 dataset

Since the proposed multi-view approach demonstrated stable and positive influence on the accuracy of the models (see Table 1), the approach was additionally tested with CIFAR-10 dataset (Krizhevsky, 2009). The dataset was manually corrupted with *asymmetric noise* in labels (Xia et al., 2021; Ma et al., 2020). Both, in the main experiments (see Table 1) and in this experiments the 9-layer CNN network presented in (Han et al., 2018; Xia et al., 2021) was used. For all experiments, the Adam optimizer with momentum of 0.9 and weight decay of 0.01 was used, the batch size was set to 128, the models were trained in 200 epochs with an initial rate of 0.001 and a schedule in correspondence to paper by Xia et al. (2021). Unlike the paper by Xia et al. (2021) the inference models correspond to the best validation loss. The models were initialized with the random weights before the training. Each experiment was performed within 5 parallel tests with random seeds of $[42, 0, 17, 9, 3]$. Unlike the main experiments of this paper, the technique of label smoothing was applied with noisy labels (Wei et al., 2021) in this series of experiments. The combination of label smoothing and multi-view predictions allowed the model to exceed some of the results of other authors (see Table 2). The number of multi-views was set to 50.

Table 2: Comparison of the proposed multi-view (MV) approach with alternative approaches to noisy data processing. CIFAR-10 dataset and its corrupted version with noise in labels were used. Top two results are highlighted.

MODEL	**LS** ratio	**#PAR.** (train. par.)	**#EP.**	**ACCURACY, %** Clean labels	20% asym. noise
		Proposed models			
9-layer CNN	0 / 0.4	4.4 M (all)	200	87.19 ± 1.03	80.67 ± 0.52
9-layer CNN + MVM	0 / 0.4	4.4 M (all)	200	$\mathbf{92.28 \pm 0.65}$	87.55 ± 0.22
9-layer CNN + MVWCo-S	0 / 0.4	4.4 M (all)	200	$\mathbf{92.34 \pm 0.66}$	$\mathbf{87.60 \pm 0.23}$
		Xia et al. (2021)			
9-layer CNN + Co-teach.	-	4.4 M (all)	200	-	83.87 ± 0.24
9-layer CNN + JoCor	-	4.4 M (all)	200	-	80.96 ± 0.25
9-layer CNN + CNLCU-S	-	4.4 M (all)	200	-	85.06 ± 0.17
		Wei et al. (2021)			
ResNet34 + LS	0 / 0.4	-	200	91.44 ± 0.16	$\mathbf{90.49 \pm 0.1}$