ylly

by Ylyl Yl

Submission date: 19-Dec-2023 01:33AM (UTC-0500)

Submission ID: 2262419173

File name: 809_01_Revised_manuscript_1219.docx (700.2K)

Word count: 3791

Character count: 20406

Decomposed Channel based Multi-Stream Ensemble: improving consistency targets in semi-supervised 2D pose estimation

Abstract

Objectives: In pose estimation, semi-supervised learning is a crucial approach to overcome the lack of information problem of labeled data. However, for semi-supervised learning, the insufficient number of labeled samples also severely affects its functionality. The fewer labeled the data, the less stable the prediction. Deep ensemble is a good way to improve model accuracy and stability. However, the training time of model ensemble is long and the resource consumption is high, so it cannot be applied in many practical scenarios. Therefore, the methods we propose the Decomposed Channel based Multi-Stream Ensemble (DCMSE) network, which can extend a single model to a stream-ensemble structure and generate the ensemble prediction to solve the large variance of prediction from the lack of labeled data, and improve the performance. The Channel Deconstruction and Ensembling (CDE) module makes the network benefits from both diversity and commonality by implementing ensemble without increasing the size of parameters. The output features are split into two parts, common-channels and private-channels. In feature sampling, on the one hand, common channels can provide commonality between streams. On the other hand, private channels can provide diversity for each stream and avoid homogenization of the predictions for each stream. Both diversity and commonality allow the network to not only gain in the ensemble of streams, but also improve the prediction accuracy of each stream itself. Results: Moreover, we propose mean-stream consistency constraints and cross-stack consistency constraints to obtain gains from unlabeled data. The Mean-Stream (MS) consistency constraint uses multi-stream ensemble prediction to additionally supervise each stream. Based on the characteristics of the Stacked Hourglass model, the Cross-Stage consistency constraint (CS) uses the forecasting results of later stages to supervise the forecasting of previous stages from the perspective of stages.Conclusion: Our approach achieves better results than SOTAs on the FLIC and Openfield-Pranav and our Sniffing data-set. Specifically, on the MSE, our method achieves at least 0.88, 0.13, and 0.08 improvements over the SOTA method on the FLIC, Openfield-Pranav, and our Sniffing datasets, respectively.

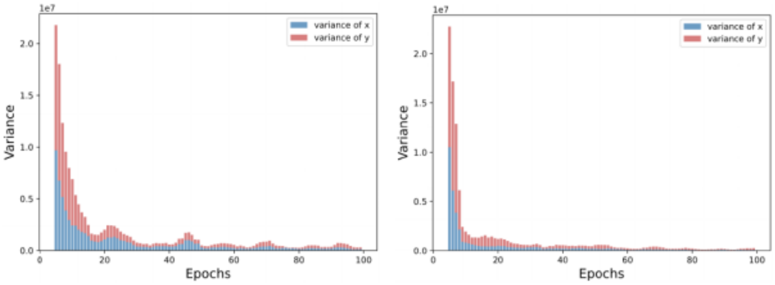
Keywords: Multi-Stream Ensemble, UDA-AP, Openfield-Pranav, DCMSE network, CDE module.

1. Introduction

Pose estimation is an essential subject in computer vision, and it has achieved many outstanding results, such as (Newell et al., 2016; Sun et al., 2019; Tompson et al., 2014; Wei et al., 2016; Xiao et al., 2018). All the above methods achieve excellent results with sufficient labeled data. However, these methods are highly sensitive to the amount of labeled data. Manual annotation of poses is complex and costly. Therefore, reducing the amount of labeled data has become a focus of research.

Currently, unsupervised pose estimation based on transfer learning has achieved exciting results. The main idea of UDA-AP (Li et al., 2021) is to gain knowledge from synthetic animal data and apply this knowledge to target domain recognition using transfer learning. However, due to the great difficulty of pose estimation, when the difference between source and target domains is large, a few of labeled data is still needed to guide the model training. Therefore, semi-supervised pose estimation based on few labeled samples still plays an essential role.

The main role of semi-supervised learning is to use a small amount of labeled data and a large amount of unlabeled data simultaneously, so that the model can achieve better predictive performance. Currently, semi-supervised learning methods are mainly applied to classification tasks. In general, there are two categories of semi-supervised learning: consistency-based and pseudo-label-based method.



(a) Variance of single model (b) Variance of ours

Figure 1. The variance of single-model and our model.

The idea of consistency-based methods (Berthelot et al., 2019; Laine and Aila, 2016; Sajjadi et al., 2016; Sohn et al., 2020; Tarvainen and Valpola, 2017) is that the model should have consistent predictions for different augmented samples of the same sample. Therefore, the main approach of such methods is to construct long-term stable and effective consistency-based supervision, thereby obtaining additional supervision from unlabeled data.

In pseudo-label-based methods (Arazo et al., 2020; Berthelot et al., 2019; Lee et al., 2013; Radosavovic et al., 2018; Wang et al., 2018; Xie et al., 2020; Zhai et al., 2019), an initial model is first trained using labeled data, and then use the initial model predicts the unlabeled data to generate pseudo-labels. Finally, using pseudo-labels to train the initial model itself. The bottleneck of this approach is that the quality of the generated pseudo-labels deeply depends on the initial model, and the noise of the pseudo-labels can easily degrade the model performance.

In the above semi-supervised methods, the model predictions are prone to fluctuation when there is less labeled data. The less labeled data, the less stable the prediction. As shown in Fig. 1, the predictions of single-model-based methods are prone to fluctuations and have large variance. This can lead to a reduction in accuracy. In extreme cases, the semi-supervised framework breaks down, as the model fails to provide valid predictions. The model collapse mentioned in (Xie et al., 2021) occurs, that is, the prediction accuracy of semi-supervised learning is lower than that of supervised one.

In this work, we propose a "no-cost" stream-based ensemble method to solve the problem of unstable prediction, in the case of few labeled data. The Channel Deconstruction and Ensembling (CDE) can sample the output features on the channel dimension, form multiple different features, and predict them through the corresponding FC layer. The output features are split into two parts, common-channels and private-channels. In feature sampling, on the one hand, the common channel can provide commonality between streams. On the other hand, the private channels can provide diversity for each stream and avoid the homogenization of the predictions of each stream. Both diversity and commonality can not only make the network gain in stream ensembling, but also improves the prediction accuracy of each stream itself.

In addition, we propose two consistency constraints to further improve the accuracy of the ensemble prediction of multiple stream. The Mean-Stream (MS) consistency constraint uses multi-stream ensemble prediction to additionally supervise each stream. Based on the characteristics of the Stacked Hourglass model, the Cross-Stage consistency constraint (CS) uses the forecasting results of later stages to supervise the forecasting of previous stages from the perspective of stages. The project code and Sniffing dataset are publicly available on https://github.com/Qi2019KB/DCMSE/tree/master.

Summary of our main contributions:

- We propose a simple structure, named Decomposed Channels based Multi-Stream Ensemble (DCMSE), to extend a model to an ensemble form easily. In DCMSE, the private channel groups of each stream can improve the diversity, which can accelerate the performance of multi-stream ensemble. Moreover, the co-channel group allows the model to learn common features from all streams, which can improve the performance of each stream itself.
- We propose the Mean-Stream Consistency Loss based on the multi-stream ensemble and the Cross-Stage Consistency Loss based on the cascade structure of the pose estimation model. Both of them allow the model to efficiently gain additional supervision from unlabeled data.
- With a few labeled data, our method achieves better results than the state-of-art pose estimation method. We implemented this result on the public data-set FLIC (Sapp and Taskar, 2013), Openfield-Pranav (Mathis et al., 2018), and our own Sniffing data-set.

2. Related Work

Semi-supervised learning. Semi-supervised learning (SSL) (Higuchi et al., 2022; Li et al., 2022; Njima et al., 2022; Xu et al., 2022) focus on reducing the need for labeled data. The main idea is to use a small amount of labeled data and a large amount of unlabeled data simultaneously, so that the model can achieve better predictive performance. In general, there are two categories of semi-supervised learning.

One popular approach is based on the pseudo-labeling (Arazo et al., 2020; Lee et al., 2013; Radosavovic et al., 2018). Labeled data is used to train initial model as per-trained model, and the model generates pseudo-labels from unlabeled data, then the model is trained with artificial labels and pseudo-labels. Another one is the consistency regularization. It encourages the model have consistent predictions for different augmented samples of the same sample.

Deep Ensemble. Training multiple deep neural networks (DNNs) is to improve the predictive performance. Wang et al., 2021 is to average the parameters stored at multiple checkpoints to obtain a new model. The advantage of this ensemble approach is that the training cost is low, but the prediction accuracy barely increases due to the lack of diversity.

Another approach is to ensemble multiple deep models and take the average prediction as the final output. This approach can improve performance, but is costly to train.

3. Method

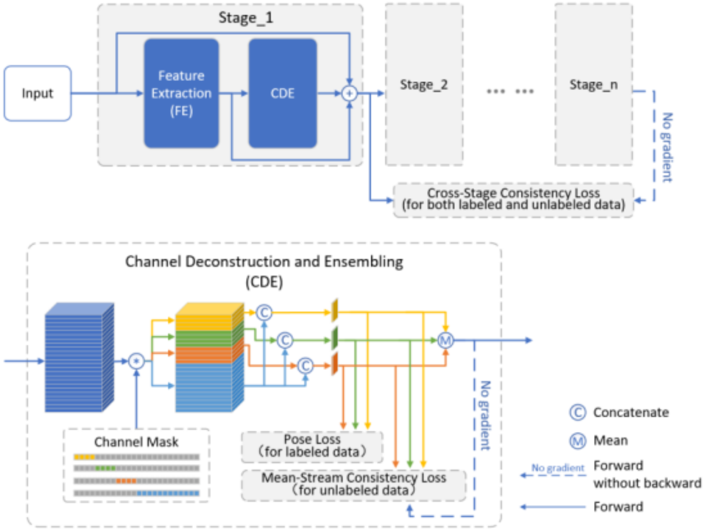

Stage_1
Input
Feature Extraction (FE)
CDE
Stage_2
Stage_n
No gradient
Cross-Stage Consistency Loss (for both labeled and unlabeled data)
Channel Deconstruction and Ensembling (CDE)
C
M
Channel Mask
Pose Loss (for labeled data)
Mean-Stream Consistency Loss (for unlabeled data)
No gradient
Concatenate
Mean
No gradient
Forward without backward
Forward

Figure 2. Overview of the proposed Decomposed Channel based Multi-Stream Ensemble (DCMSE) network.

3.1 2D Pose Estimation with Multi-Stage Model

In 2D Pose estimation, the main task is to predict the accurate position of K key points which are predefined manually in the input Image $\mathbf{I}$. Most recent approaches transform this problem into predicting K Gaussian heat-maps $\mathbf{H}$ that encode the probability of each key point at each spatial location in $\mathbf{I}$. Therefore, a Gaussian kernel function is used to transform the coordinate values into Gaussian heat maps, where one key point corresponds to one heat-map, and the heat maps are used as ground-truth for model training. During prediction, the location with the highest response value is selected from the heat-map generated by the model as the predicted location of the key point. Denote the labeled training data-set as $\mathcal{L} = \{(\mathbf{I}^l, \mathbf{H}^l)\}_{l=1}^{N^L}$. To train the 2D pose estimation model $\mathbf{f}(\cdot, \theta)$, we write simply $\mathbf{f}(\cdot)$. we minimize the MSE loss between the prediction and ground-truth heat map. Therefore, the Pose loss, as L_{pose}, can be expressed as:

$$L_{pose} = \mathbb{E}_{\mathbf{I}\in\mathcal{L}} \parallel \mathbf{f}(\mathbf{I}) - \mathbf{H} \parallel^2, \qquad (1)$$

The M stage pose estimation model $\mathbf{g}(\cdot, \theta)$, we write simply $\mathbf{g}(\cdot)$, such as Stacked Hourglass (Newell et al., 2016), stacks multiple modules with the same structure end to end, and feeds the previous model's predictions and features to the next. There is supervised learning at each stage, called intermediate supervision, to ensure performance at each stage. So, for image $\mathbf{I}$, the input $\mathbf{X}_m^{\mathbf{I}}$ of the m-th stage can be expressed as:

$$\mathbf{X}_m^{\mathbf{I}} = \begin{cases} \mathbf{I} & m = 1 \\ \mathbf{I} + P_{m-1}^{\mathbf{I}} + \varphi_{m-1}^{\mathbf{I}} & otherwise, \end{cases} \qquad (2)$$

where $P_{m-1}^{\mathbf{I}} = \mathbf{g}_{m-1}(\mathbf{X}_{m-1}^{\mathbf{I}})$ is the prediction in the $(m-1)$-th stage of the model $\mathbf{g}(\cdot)$, $\varphi_{m-1}^{\mathbf{I}}$ is the feature extracted by the feature extraction module in the $(m-1)$-th stage of the model $\mathbf{g}(\cdot)$.

3.2 Channel Deconstruction and Ensembling

In each stage of the multi-stage network, the Channel Deconstruction and Ensembling (CDE) module is located behind the feature extraction (FE) module. The CDE deconstructs the features extracted by FE in the channel dimension and forms n new features corresponding to n streams. We assume that the feature extracted from a sample by the FE is of $C \times H \times W$ dimension, where H, W, C denote the height, width and number of channels, respectively. Specifically, the CDE extracts n features with channel C' from the C-channels in the input feature according to the splitting factor α as the private features of streams. The feature corresponding to the remaining channels is treated as common feature and shared by all streams. Each private feature is concatenated with the common feature and passed to the regressor of each stream for prediction. In the m-th stage, consider the ensemble output $\widetilde{\mathbf{P}}^{\mathbf{I}}_m$ of the image $\mathbf{I}$ by averaging each stream's prediction $\mathbf{P}^{\mathbf{I}}_{m,n}$, i.e.,

$$\widetilde{\mathbf{P}}^{\mathbf{I}}_m = \frac{1}{n}\sum_{n=1}^{N} \mathbf{P}^{\mathbf{I}}_{m,n}, \qquad (3)$$

At the same time, we pass the full features extracted by feature extraction module with the ensemble output to the next stage. So, for image $\mathbf{I}$, the input $\mathbf{X}^{\mathbf{I}}_m$ of the m-th stage can be rewritten as:

$$\mathbf{X}^{\mathbf{I}}_m = \begin{cases} \mathbf{I} & m = 1 \\ \mathbf{I} + \widetilde{P}^{\mathbf{I}}_{m-1} + \varphi^{\mathbf{I}}_{m-1} & otherwise, \end{cases} \qquad (4)$$

3.3 Mean-Stream Consistency Constraint

The M stage pose estimation model $\mathbf{g}(\cdot)$ also learns about unlabeled data. Denote the unlabeled training data-set as $\mathcal{U} = \{\mathbf{I}^u\}_{u=1}^{N^U}$.

Several studies (Ren et al., 2016; Zhang and Suganthan, 2017) have demonstrated that the performance of a single model is inferior than that of an ensemble of models. Therefore, in each stage, we minimize the MSE loss between the ensemble output and each stream's prediction. Denote the prediction of the n-th stream to sample $\mathbf{I}, \mathbf{I} \in \mathcal{U}$ in the m-th stage as $\mathbf{P}_{m,n}^{\mathbf{I}}$, the ensemble output to sample $\mathbf{I}, \mathbf{I} \in \mathcal{U}$ in the m-th stage as $\widetilde{\mathbf{P}}_{m}^{\mathbf{I}}$. Therefore, the mean-stream consistency loss, as L_{ms}, can be expressed as:

$$L_{ms} = \frac{1}{m}\frac{1}{n}\sum_{m=1}^{M}\sum_{n=1}^{N}\mathbb{E}_{\mathbf{I}\in\mathcal{U}} \parallel \mathbf{P}_{m,n}^{\mathbf{I}} - \widetilde{\mathbf{P}}_{m}^{\mathbf{I}} \parallel^{2}, \tag{5}$$

Note that the consistency constraint affects both parties. To avoid a negative effect on the model performance, we compute this loss using a replica of the ensemble output with the gradient information removed, so that the model is not updated from the ensemble output.

3.4 Cross-Stage Consistency Constraint

As stated in Newell et al., 2016, in multi-stage pose estimation model, subsequent stages allow the high-level features, extracted by previous stage, to be processed again to make the network to best refine predictions. Moreover, the more accurate the prediction at the front stage, the more accurate the prediction at the back stage, which is in line with the goal of intermediate supervision. Therefore, we minimize the MSE loss between the ensemble output of the first stage and the ensemble output of the last stage. Therefore, the cross-stage consistency loss, as L_{cs}, can be expressed as:

$$L_{cs} = \mathbb{E}_{\mathbf{I}\in(\mathcal{L}\cup\mathcal{U})} \parallel \widetilde{\mathbf{P}}_{1}^{\mathbf{I}} - \widetilde{\mathbf{P}}_{m}^{\mathbf{I}} \parallel^{2}, \tag{6}$$

Note that the cross-stage consistency constraint can work on both labeled and unlabeled data. For the same reason as in L_{ms}, we compute this loss using a replica of the final-stage ensemble output with the gradient information removed.

3.5 Total Loss

We summarize the objective function as follows:

$$L = \lambda_{pose}L_{pose} + \lambda_{ms}L_{ms} + \lambda_{cs}L_{cs}, \tag{7}$$

where λ_{pose} , λ_{ms}, λ_{cs} are the weights to balance all losses.

4. Experiment

We evaluate our model on three datasets. FLIC (Sapp and Taskar, 2013) data-set, an public human pose data-set; Openfield-Pranav (Mathis et al., 2018) data-set, an public mouse data-set; Sniffing data-set, a mouse data-set collected by ourselves.

4.1 Implementation Details

We use Stacked Hourglass (Newell et al., 2016) with a stack number of 3 as the pose estimation model. When the CDE is created, the number of streams, N, is 3 and the splitting factor, α, is 0.2.

All input images are resized to 256×256 pixels. And we use the data augmentation that includes random rotation $(+/-30$ degrees$)$, and random scaling $(0.75 - 1.25)$ and random horizontal flip.

The initial value of λ_{pose} is the constant 10. The initial value of λ_{ms} is 0 and after 50 epochs rises to a maximum of 50. The initial value of λ_{cs} is 0 and after 50 epochs rises to a maximum of 5.

4.2 FLIC Dataset

The FLIC (Sapp and Taskar, 2013) dataset is a publicly available human pose dataset consisting of 5003 images taken from movies, including 3987 training data and 1016 test data.

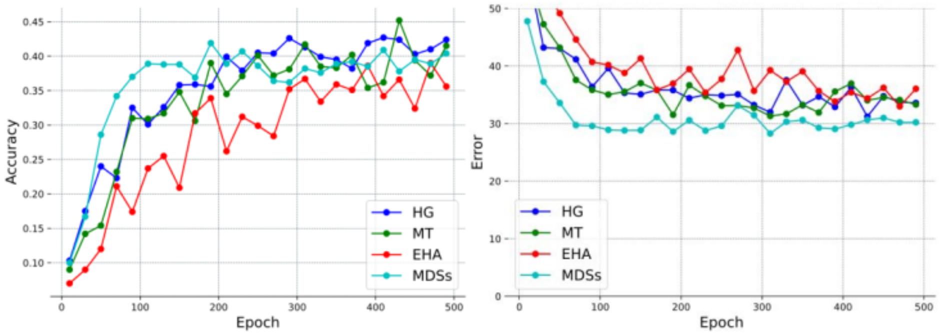


Figure 3. Comparison to the baselines with 500 training samples with 30% labeled data on the FLIC dataset. Supervised learning is training with Newell et al., 2016. And, DNCL (Shi et al., 2018), Mean Teacher (Tarvainen and Valpola, 2017), ESCP (Xie et al., 2021) and ours are semi-supervised learning.

The MSE and PCK@0.2 are reported (see results in Fig. 3 and Table. 1).

*Table 1. Comparison to the baselines on the FLIC dataset. The 500*0.3 means that there are 500 samples in the training set, of which 30% are labeled samples and 70% are unlabeled samples.*

Method	**500*0.3**		**500*0.5**	
	MSE	**PCK@0.2**	**MSE**	**PCK@0.2**
HG (Newell et al., 2016)	37.27	0.436	33.36	0.516
DNCL (Shi et al., 2018)	35.56	0.456	30.22	0.535
MT (Tarvainen and Valpola, 2017)	35.58	0.448	28.68	0.551
ESCP (Xie et al., 2021)	39.57	0.390	35.06	0.486
Ours	**32.26**	**0.490**	**27.80**	**0.574**

4.3 Openfield-Pranav Data-set

The Openfield-Pranav data-set is a publicly available mouse data-set consisting of 1000 images with one instance per image.

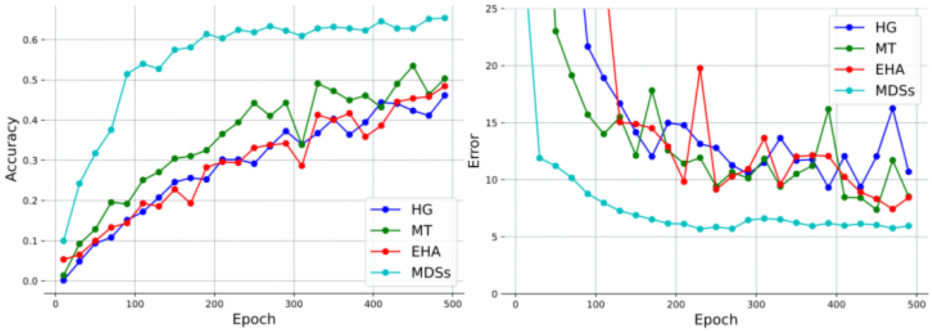


Figure 4. Error and accuracy curves for each baseline with 100 training samples with 30% labeled data on the Openfield-Pranav dataset.

The MSE and PCK@0.2 are reported (see results in Fig. 4 and Table. 2). Our method achieves the best results on the Openfield-Pranav dataset in both cases 100*0.3 and 100*0.5, and the increase amount is 1.8% and 0.7% respectively.

*Table 2. Comparison to the baselines on the Openfield-Pranav dataset. 100*0.3 means that there are 100 training samples and 30% of them are labeled.*

Method	**100*0.3**		**100*0.5**	
	MSE	**PCK@0.2**	**MSE**	**PCK@0.2**
HG	8.54	0.565	4.34	0.725
DNCL	5.56	0.647	4.13	0.756
MT	5.89	0.635	4.36	0.739
ESCP	5.87	0.617	4.25	0.717
Ours	**5.50**	**0.665**	**4.00**	**0.763**

4.4 Sniffing Dataset

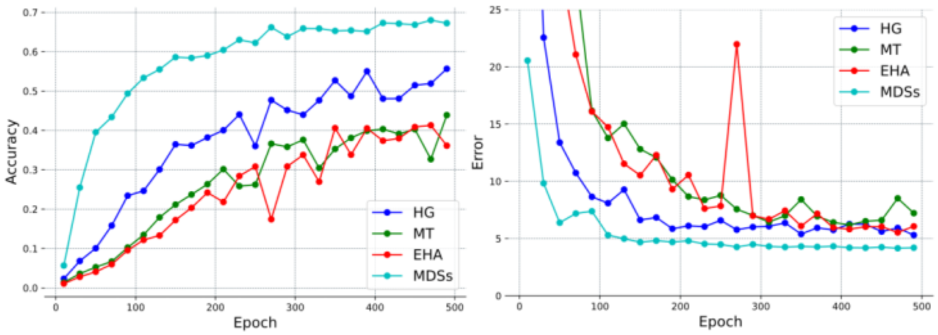


Figure 5. Error and accuracy curves for each baseline with 100 training samples with 30% labeled data on the Sniffing dataset.

The Sniffing dataset was captured by ourselves from real experimental environments. The MSE and PCK@0.2 are reported (see results in Fig. 5 and Table. 3). Our method achieves the best results on the Sniffing dataset in both cases 100*0.3 and 100*0.5, and the increase amount is 0.9% and 0.3% respectively.

*Table 3. Comparison to the baselines on our Sniffing Data-set. 100*0.3 means that there are 100 training samples and 30% of them are labeled.*

Method	**100*0.3**		**100*0.5**	
	MSE	**PCK@0.2**	**MSE**	**PCK@0.2**
HG	5.18	0.602	4.26	0.663
DNCL	4.47	0.634	3.89	0.698
MT	4.73	0.613	4.14	0.676
ESCP	5.03	0.596	3.93	0.675
Ours	**4.26**	**0.643**	**3.81**	**0.701**

5. Ablation Study

We perform ablation studies on the FLIC and Openfield-Pranav datasets.

In the CDE ablation experiments, we evaluate the role of CDE in terms of ensembling. We construct two simple models. HG-Ensemble is a simple parallel ensembling of the Stacked Hourglass (HG), taking the average predicted values of three HG as the output of the entire model. In addition, we remove the CDE module from our network (stream count is 3), which means that the regressors of all three streams share the same features as the input. Finally, we use Stacked Hourglass for supervised learning and evaluate their predictive performance.

Table 4. The result of Openfield-Pranav dataset. The CDE is evaluated from an ensemble perspective. Stream Predictions is a list of the prediction performance of all streams.

Method	**Stream Count**	**MSE**	**PCK@0.2**	**Stream's Predictions**	
				MSE	**PCK@0.2**
Supervised	-	6.78	0.571	-	-
HG-Ensemble	3	6.34	0.600	7.30, 6.99, 8.74	0.569, 0.574, 0.553
Ours (without CDE)	3	**5.89**	**0.637**	5.94, 5.92, 5.95	0.626, 0.628, 0.625

As shown in Table. 4, our network achieved significant performance gains despite the removal of CDE. This is because the prediction accuracy of the regressors for each stream in our network is significantly improved compared to HG-Ensemble. Multiple streams share the same feature, so that the feature extraction module in the network can learn to extract the common feature of higher dimensions, which can improve the accuracy of each stream's regress.

We also performed ablation experiments on the Mean-Stream(MS) and Cross-Stage (CS) consistency constraint, shown in Table. 5.

Table 5. The ablation experiments on the Mean-Stream (MS) and Cross-Stage (CS) consistency constraint. "Y" and "N" mean "using" and "not using", respectively.

Method	Dataset	MS	CS	MSE	PCK@0.2
Ours	FLIC 100*0.3	N	N	45.91	0.284
		Y	N	45.47	0.288
		Y	Y	**44.42**	**0.297**
	Openfield-Pranav 100*0.3	N	N	6.28	0.643
		Y	N	5.88	0.651
		Y	Y	**5.68**	**0.665**

6. Discussions

After completing the above experiments and proving the effectiveness of our method, we discuss two important parameters of the CDE module: the stream count N and the splitting factor α . We put different values on each of these variables and see what they do. The experimental results are shown in Table. 6.

Table 6. Compare the influence of different values of stream count N *and splitting factor* α *on the model performance.*

Method	Stream Count	Splitting Factor	MSE	PCK@0.2	Stream's Predictions	
					MSE	PCK@0.2
Ours	3	0.1	6.22	0.629	6.28, 6.26, 6.25	0.618, 0.623, 0.622
	3	0.15	5.78	0.637	5.82, 5.84, 5.84	0.632, 0.627, 0.634
	3	0.2	**5.78**	**0.648**	5.78, 5.90, 5.83	0.641, 0.644, 0.638

Ours	2	0.2	6.41	0.630	6.45, 6.45	0.620, 0.623
	3	0.2	**5.85**	**0.657**	5.73, 5.98, 6.00	0.640, 0.643, 0.648
	4	0.2	5.92	0.645	6.00, 6.00, 6.00, 6.01	0.630, 0.629, 0.634, 0.631

First, we fix the stream count N to 3 and set the splitting factor α to 0.1, 0.15 and 0.2, respectively. It can be seen that the prediction performance of the model gradually improves as the splitting factor increases.This indicates that more private features contribute to the increase of the ensembling performance.

Then, we fix the splitting factor α to 0.2 and set the stream count N to 2, 3 and 4, respectively. The results show that more streams are not always better. This is because when the number of streams increases, the size of the common features decreases, which adversely affects the model. Therefore, we need to balance these two parameters. In all experiments, we used this optimal set of parameters.

7. Conclusion

The DCMSE offers a practical and efficient solution to enhance the accuracy and stability of semi-supervised learning when labeled data is scarce. Its adaptability and superior performance make it a promising tool for a wide range of real world applications. Semi-supervised learning is a crucial approach to address the absence of labeled data, the insufficient number of labeled samples also severely affects its functionality. Model ensembles are a useful way to improve model accuracy and stability. However, model ensembles suffer from lengthy training times and extreme resource consumption, so they cannot be applied in many real-world scenarios. The main aim of our study is to build an ensemble prediction framework at a lower cost and use ensemble prediction to address inaccuracies and instabilities in semi-supervised learning when labeled data is insufficient.

The DCMSE network transforms the traditional model ensemble idea into a flow ensemble inside the model. On the one hand, it creates the multi-branch structure needed for ensemble prediction. On the other hand, it does not increase the number of model parameters due to the feature sample-based ensemble of streams. Therefore, it is easy to generalize to different tasks and to different models. With few labeled data, our method achieves better results than state-of-the-art pose estimation methods. We implement this result on the public datasets FLIC, Openfield-Pranav, and our own Sniffing dataset. In terms of MSE, our method achieves at least 0.88, 0.13 and 0.08 improvement over the SOTA method on FLIC, Openfield-Pranav and our Sniffing dataset, respectively.

Conflict of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Consent to Publish

The participant has consented to the submission in this Journal.

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Fund statement

There is no funding for this research.

Author’s contributions

Jiaqi Wu made the primary contribution to this study, which included the design and methodology of this study, the evaluation of the results, and the writing of the manuscript. Junbiao Pang and Qingming Huang contributed to the methodology of this study.

ylly

ORIGINALITY REPORT

14%	11%	10%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	5%
2	www.researchsquare.com Internet Source	3%
3	"Pattern Recognition and Computer Vision", Springer Science and Business Media LLC, 2021 Publication	1%
4	"Computer Vision – ECCV 2016", Springer Science and Business Media LLC, 2016 Publication	1%
5	"Computer Vision – ECCV 2020", Springer Science and Business Media LLC, 2020 Publication	<1%
6	Shulei Wang, Katsunori Mizuno, Shigeru Tabeta, Kei Terayama et al. "An efficient segmentation method based on semi-supervised learning for seafloor monitoring in Pujada Bay, Philippines", Ecological Informatics, 2023 Publication	<1%

7	Tao Tao, Kan Liu, Li Wang, Haiying Wu. "Image Recognition and Analysis of Intrauterine Residues Based on Deep Learning and Semi-Supervised Learning", IEEE Access, 2020 Publication	<1 %
8	www.mdpi.com Internet Source	<1 %
9	Yue Duan, Zhen Zhao, Lei Qi, Lei Wang, Luping Zhou, Yinghuan Shi, Yang Gao. "MutexMatch: Semi-Supervised Learning With Mutex-Based Consistency Regularization", IEEE Transactions on Neural Networks and Learning Systems, 2022 Publication	<1 %
10	www.ncbi.nlm.nih.gov Internet Source	<1 %
11	export.arxiv.org Internet Source	<1 %
12	vtechworks.lib.vt.edu Internet Source	<1 %
13	eprint.iacr.org Internet Source	<1 %
14	hdl.handle.net Internet Source	<1 %

15	Chetan L. Srinidhi, Seung Wook Kim, Fu-Der Chen, Anne L. Martel. "Self-supervised driven consistency training for annotation efficient histopathology image analysis", Medical Image Analysis, 2021 Publication	<1 %
16	Lei Tian, Peng Wang, Guoqiang Liang, Chunhua Shen. "An Adversarial Human Pose Estimation Network Injected with Graph Structure", Pattern Recognition, 2021 Publication	<1 %
17	Chen Jun, Jiang Lan. "Research on Optimization and Risk Control of E-commerce Supply Chain Logistics Management Mode Based on the Internet of Things and Blockchain Technology", Research Square Platform LLC, 2023 Publication	<1 %
18	Chetan L. Srinidhi, Seung Wook Kim, Fu-Der Chen, Anne L. Martel. "Self-supervised driven consistency training for annotation efficient histopathology image analysis", Medical Image Analysis, 2022 Publication	<1 %
19	Jingwei Li, Yuan Li, Jie Tan, Chengbao Liu. "Bridging the gap with grad: Integrating active learning into semi-supervised domain generalization", Neural Networks, 2024	<1 %

Publication

20	Yo-Seb Jeon, Jun Li, Nima Tavangaran, H. Vincent Poor. "Data-Aided Channel Estimator for MIMO Systems via Reinforcement Learning", ICC 2020 - 2020 IEEE International Conference on Communications (ICC), 2020 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		