Final_project_Report_1002

April 18, 2024

1 Title: Facial Expression Recognition using Temporal Ensembling for Semi-Supervised Learning

1.0.1 Group Member Names :

Athuri Vivekanantharasa 200583830

Milap Rajeskumar Milap 200579676

1.0.2 INTRODUCTION:

The existing research conducted a comprehensive study on semi-supervised learning methods for Facial Expression Recognition (FER). They explored eight semi-supervised learning methods, including Pi-Model, Pseudolabel, Mean-Teacher, VAT, MixMatch, ReMixMatch, UDA, and FixMatch. The study evaluated these methods on three FER datasets: FER13, RAF-DB, and AffectNet. Their analysis focused on comparing the performance of these methods against fully-supervised training, highlighting the potential of semi-supervised approaches to achieve comparable results with smaller labeled datasets. The research aimed to reduce the reliance on large annotated datasets for training deep neural networks in FER tasks.

In our study, we focuses on enhancing the performance of facial expression recognition systems through the implementation of temporal ensembling within a semi-supervised learning framework. By leveraging temporal dynamics in the learning process, we aim to reduce the reliance on annotated data, thereby addressing a key challenge in the field. We conducted experiments on the FER2013 dataset exploring four distinct scenarios with varying numbers of labeled datasets per class. Through our empirical investigation, we aim to elucidate the effectiveness of temporal ensembling in semi-supervised facial expression recognition, paving the way for more accurate and reliable systems in affective computing and human-machine interaction domains.

1.0.3 AIM :

The research aimed to reduce the reliance on large annotated datasets for training deep neural networks in FER tasks.

1.0.4 Github Repo:

https://github.com/Athuriviveka/FER/tree/main

1.0.5 DESCRIPTION OF PAPER:

This paper provides a comprehensive exploration of semi-supervised learning methodologies within the domain of Facial Expression Recognition (FER). Conducted by Shuvendu Roy and Ali Etemad from Queen's University, Canada, the study investigates eight prominent semi-supervised techniques, including Pi-Model, Pseudolabel, Mean-Teacher, VAT, MixMatch, ReMixMatch, UDA, and FixMatch. Through rigorous evaluation on three benchmark FER datasets—FER13, RAF-DB, and AffectNet—the research showcases the potential of semi-supervised methods to enhance FER accuracy while minimizing the need for extensive human annotation. Additionally, a sensitivity analysis uncovers key hyperparameters' impact on method efficacy, offering valuable insights for optimal configuration in FER tasks.

Our contribution aims to reduce reliance on annotated data for facial expression recognition by introducing a novel implementation of the semi-supervised method temporal ensembling into the FER domain. Temporal ensembling leverages temporal dynamics within the learning process, aiming to capture and exploit sequential patterns in facial expressions to enhance recognition accuracy and robustness. Specifically, we evaluate the performance of temporal ensembling on the FER2013 dataset across four distinct scenarios, varying the number of labeled datasets per class—10, 25, 100, and 250. Through this empirical exploration, our contribution seeks to elucidate the effectiveness of temporal ensembling in semi-supervised FER. By addressing these research questions, we aim to contribute to the ongoing pursuit of reducing reliance on annotated data for facial expression recognition, thereby enriching the landscape of affective computing and human-machine interaction technologies.

1.0.6 PROBLEM STATEMENT :

Facial Expression Recognition (FER) systems heavily rely on large annotated datasets for training deep neural networks. However, the collection of such datasets is laborious and costly. Despite recent advancements in semi-supervised learning methods, their application in FER remains relatively unexplored. Existing research demonstrates the potential of semi-supervised approaches to reduce the reliance on large annotated datasets. However, there is a need to evaluate and extend these methods to improve FER accuracy and reduce annotation costs further.

1.0.7 CONTEXT OF THE PROBLEM:

FER systems have promising applications in various domains, including human-computer interaction, healthcare, and multimedia analysis. However, challenges such as data annotation costs, dataset size, and model robustness hinder their widespread adoption. Existing research has shown that semi-supervised learning methods offer a viable solution to these challenges by leveraging unlabeled data to enhance model performance. By reducing the need for large annotated datasets, semi-supervised methods have the potential to democratize FER technology, making it more accessible and cost-effective.

1.0.8 SOLUTION:

Building upon existing research, our contribution introduces a novel implementation of the semi-supervised method temporal ensembling into the FER domain. Temporal ensembling leverages temporal dynamics within the learning process, aiming to capture and exploit sequential patterns in facial expressions to enhance recognition accuracy and robustness. Specifically, we evaluate the performance of temporal ensembling on the FER2013 dataset across four distinct scenarios, varying the number of labeled datasets per class—10, 25, 100, and 250. Through this empirical exploration, our contribution seeks to elucidate the effectiveness of temporal ensembling in semi-supervised FER. By deducing reliance on annotation data, our aim is to advance the development of more cost-effective and accessible facial expression recognition systems, thereby enriching the landscape of affective computing and human-machine interaction technologies.

2 Background

Reference	Explanation	Dataset/Input	Weakness
https://github.com/ShuvenduRoy/SSL_FER/tree/main	research by has shown the potential of semi-supervised learning methods to reduce reliance on large annotated datasets for Facial Expression Recognition (FER). These methods leverage unlabeled data to enhance model performance and reduce annotation costs.	FER datasets (e.g., FER13, RAF-DB, AffectNet)	its neglect of temporal dynamics in facial expressions. This oversight limits the ability to capture sequential patterns, which could enhance recognition accuracy.

---

3 Implement paper code :

Our implementation of temporal ensembling in the context of Facial Expression Recognition (FER) involved several key steps aimed at leveraging temporal dynamics to enhance recognition accuracy and reduce reliance on annotated data. Here's a brief overview of our approach:

1. Model Selection

We selected a suitable deep learning architecture for FER tasks, considering its compatibility with temporal ensembling techniques.

2. Data Preprocessing

The FER2013 dataset underwent preprocessing steps to ensure compatibility with our chosen model and temporal ensembling approach. Splitting into Labeled and Unlabeled Data:

We partitioned the training data into labeled subsets containing 10, 25, 100, and 250 labeled samples per class, respectively, and an unlabeled subset.

3. Integration of Temporal Ensembling

We implemented the temporal ensembling algorithm into the training pipeline, allowing the model to exploit temporal consistency in facial expressions over time.

4. Semi-Supervised Learning Strategy

The model was trained using a semi-supervised learning strategy, where labeled data subsets were used for supervised training, and the unlabeled subset was used for unsupervised learning and consistency regularization.

5. Training Procedure

The training procedure involved iteratively updating the model parameters using both labeled and unlabeled data, guided by temporal consistency constraints enforced by temporal ensembling.

6. Evaluation Procedure

We evaluated the trained model on the FER2013 dataset under the four distinct scenarios defined by the varying number of labeled samples per class. Performance Analysis:

We conducted a comprehensive analysis of the model's performance, focusing on recognition accuracy, robustness, and convergence behavior across different labeling scenarios.

3.0.1 Contribution Code :

Our contribution code encompassed the following key components:

1. Data Splitting Functions

Implementation of functions to partition the training data into labeled subsets with varying numbers of samples per class and an unlabeled subset.

2. Temporal Ensembling Algorithm:

Implementation of the temporal ensembling algorithm, including mechanisms for pseudo-labeling and consistency regularization.

3. Semi-Supervised Training Pipeline:

Code snippets illustrating the semi-supervised training pipeline, highlighting the integration of temporal ensembling and the handling of labeled and unlabeled data.

4. Evaluation Metrics Calculation:

we evaluated the performance of our models using key metrics such as train and test loss, accuracy, confusion matrix, classification report, and visualized sample predictions. These metrics allowed us to assess the models' accuracy, robustness, and generalization across different scenarios with varying amounts of labeled data.

3.0.2 Results :

The results obtained from our implementation of temporal ensembling for semi-supervised FER on the FER2013 dataset are presented below.

Scenario	Precision	Recall	F1-Score	Accuracy	Support
10 Labeled	0.12	0.10	0.11	0.14	958
25 Labeled	0.12	0.13	0.13	0.16	958
100 Labeled	0.15	0.13	0.13	0.14	958
250 Labeled	0.15	0.16	0.15	0.16	958

3.0.3 Observations :

Upon training the model on the FER2013 dataset using temporal ensembling techniques across four scenarios with varying numbers of labeled datasets per class (10, 25, 100, and 250), the following observations were made:

1. Train Accuracy Reached 1 After 3 Epochs: The model achieved perfect accuracy on the training data after just three epochs, indicating overfitting.
2. Test Accuracy Remained Around 0.4 Across 60 Epochs: Despite the high train accuracy, the model's performance on unseen data (test accuracy) plateaued around 0.4, suggesting poor generalization.
3. Train Loss Decreased While Test Loss Increased: The train loss consistently decreased with each epoch, indicating that the model was fitting the training data increasingly well. However, the test loss increased over time, indicating a decline in performance on unseen data, likely due to overfitting.

These observations highlight the challenge of achieving good generalization performance with temporal ensembling on the FER2013 dataset and underscore the need for further investigation into regularization techniques or the acquisition of additional data to mitigate overfitting.

3.0.4 Conclusion and Future Direction

3.0.5 Learnings :

Through the implementation of temporal ensembling techniques on the FER2013 dataset, several valuable insights have been gained:

1. Understanding the challenges of semi-supervised learning in the facial expression recognition domain.
2. Identifying the limitations of temporal ensembling in capturing complex temporal dynamics of facial expressions.

3.0.6 Results Discussion :

The results obtained from our experiments shed light on the performance of temporal ensembling in semi-supervised FER. While the model demonstrated high train accuracy, its generalization performance on unseen data remained limited, indicating the presence of overfitting. This highlights the importance of addressing overfitting issues and improving the model's ability to generalize to unseen data.

3.0.7 Limitations :

Several limitations were encountered during our research:

Limited generalization performance on the test set due to overfitting. Challenges in capturing subtle temporal dynamics in facial expressions with temporal ensembling. Dependency on the quality and quantity of labeled data for effective semi-supervised learning

---

3.0.8 Future Extension :

To address the limitations and further advance the research in this area, several future directions can be explored:

Investigation of advanced regularization techniques to mitigate overfitting, such as dropout and weight decay. Exploration of alternative semi-supervised learning approaches that can better leverage temporal dynamics in facial expressions. Acquisition of additional labeled data to improve the model's generalization performance and robustness. Integration of domain-specific knowledge or pre-trained representations to enhance feature extraction and representation learning. By addressing these limitations and exploring future extensions, we aim to contribute to the development of more accurate and reliable facial expression recognition systems, thereby enriching the landscape of affective computing and human-machine interaction technologies.

4 References:

[1]: Freitas, L. (n.d.). Tensorfreitas/Temporal-Ensembling-for-Semi-Supervised-Learning. GitHub. Retrieved from https://github.com/tensorfreitas/Temporal-Ensembling-for-Semi-Supervised-Learning

Ferretj. (n.d.). Ferretj/Temporal-Ensembling. GitHub. Retrieved from https://github.com/ferretj/temporal-ensembling

Hiram64. (n.d.). Hiram64/Temporal-Ensembling-Semi-Supervised. GitHub. Retrieved from https://github.com/hiram64/temporal-ensembling-semi-supervised

Ferretj. (2018, January 22). Temporal ensembling for semi-supervised. Retrieved from https://ferretj.github.io/ml/2018/01/22/temporal-ensembling.html

Ferretj. (n.d.). Temporal ensembling for semi-supervised. Papers with Code. Retrieved from https://paperswithcode.com/paper/temporal-ensembling-for-semi-supervised