Memory Integrity of CNNs for Cross-Dataset Facial Expression Recognition

Dylan C. Tannugi and Alessandro L. Koerich
Department of Software and IT Engineering
École de Technologie Supérieure (ÉTS)
Montréal, QC, Canada.

Alceu S. Britto Jr.
Post-graduate Program in Informatics (PPGIa)
Pontifical Catholic University of Paraná (PUCPR)
Curitiba, PR, Brazil.

Introduction

- **Contextualization**
  - For many applications **deep neural networks** are the state of the art.
  - This is the case of **CNN** for Facial Expression Recognition (**FER**).

  - **FER** is very important for **human-computer interaction**, **emotion analysis**, **detection of driver fatigue**, to cite a few possible applications.
  - **Difficulties**: lack of data for training and labeling data is not trivial.
  - **Alternative**: Domain adaptation – transfer learning and cross-datasets.

Introduction

- **Problem Definition**
  - Transfer learning strategies may not preserve the CNN memory integrity.
  - Adapting an existing CNN to a new dataset may lead it to forget the source dataset.

- **Main Research Questions**
  - What is the impact of different transfer learning strategies in the CNN memory integrity using cross-dataset?

  - Which action leads to the best performance on both source and target datasets?
    - Should we fully train the model?
    - Should we rebuild it from scratch or preserve the already learned parameters?
    - Should we just slightly adapt it to the new data?

  - In fact, whether transfer learning is useful or not?

State of the Art

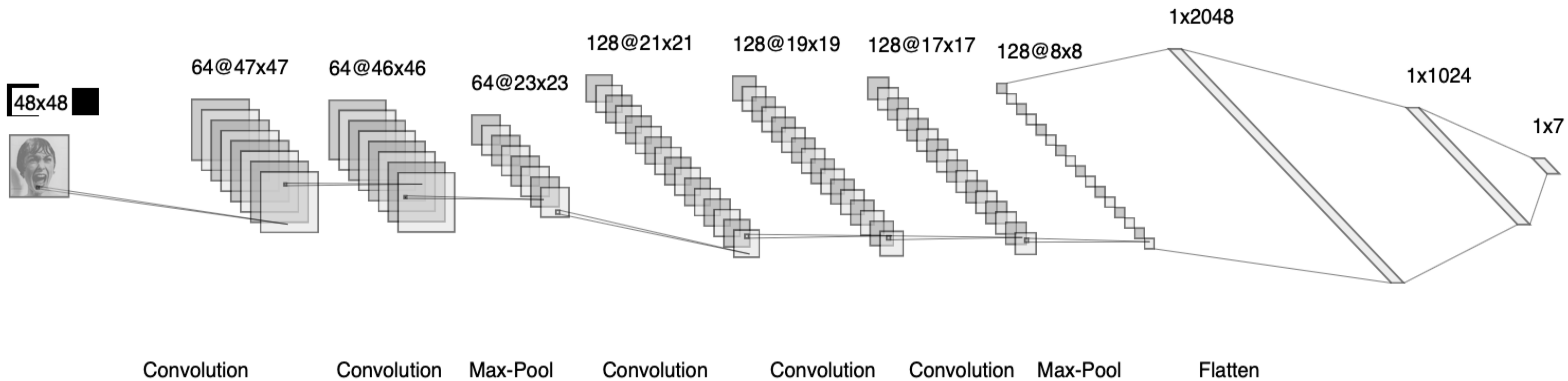
- **Observation**
  - Most of the works focus on adapting an existing CNN to a new dataset forgetting the performance on the source dataset.

- **Literature Reviews:**
  - Li and Deng et al. (2018) - a comprehensive survey on deep facial expression recognition.
  - Zhang et al (2019) - a review of transfer-learning methods for cross-dataset.

- **Interesting Contributions on Domain adaptation (to cite a few)**
  - Mayer et al. (2014) - facial expression recognition that is evaluated on multiple datasets.
  - Mousavi et al. (2016) - CNN trained on the Cohn-Kanade (CK) dataset and tested on same dataset and on the JAFFE dataset.
  - Zavarez et al. (2017) - CNN with two different initialization methods: randomly initialized weights and weights from a pre-trained VGG-Face mode.
  - Wang et al. (2018) - an unsupervised domain adaptation method, especially suitable for small unlabeled target datasets (using GAN).

Method: Deep model

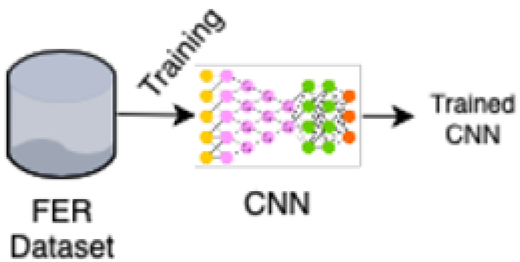
- **Architecture of the proposed CNN**



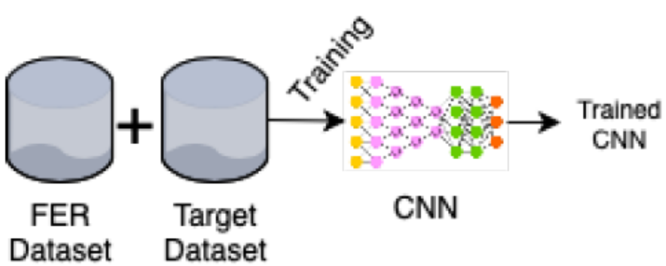
- **Simplified version of a VGG network**
  - Less convolutional layers when compared to the VGG model.
  - Number of parameters: 19.2M (7 times less than a VGG).
  - Input: gray level image of 48x48 pixels.
  - Drop-out of 50% used after the two-first fully connected layers.

Method: Training Strategies

- **Training on the source dataset**
  - FER dataset is divided into training (80%), validation (10%) and testing (10%) sets.
  - Use to generate the baseline (BL) models.

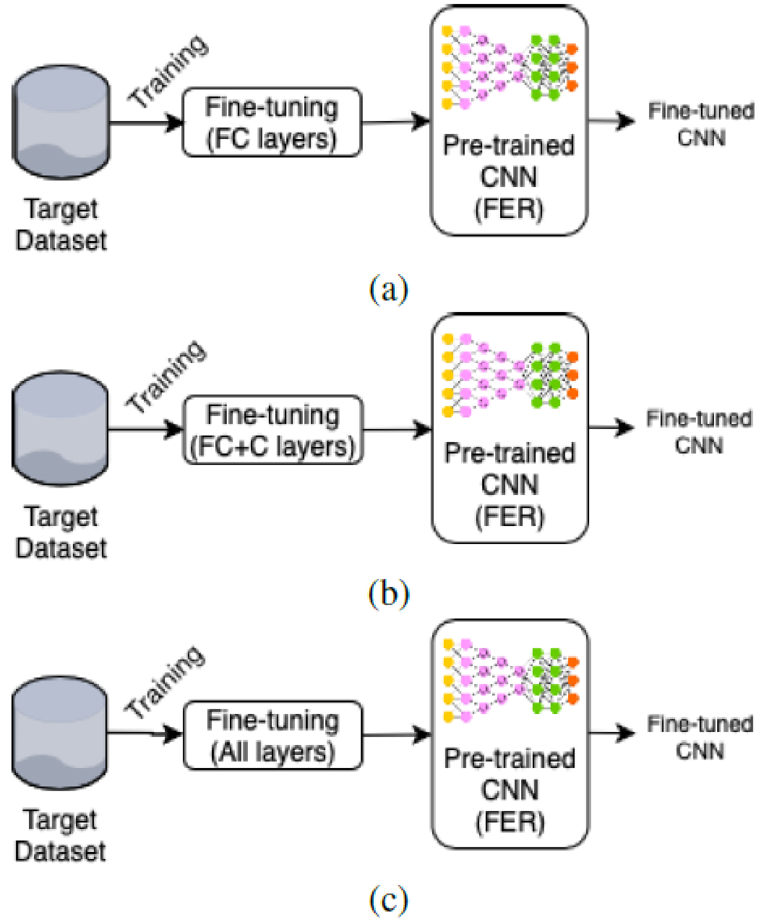


- **Training on the source + target datasets (merge)**
  - Resulting dataset is divided into training (80%), validation (10%) and testing (10%) sets.
  - Use to generate our FU models (FU = fusion of datasets).



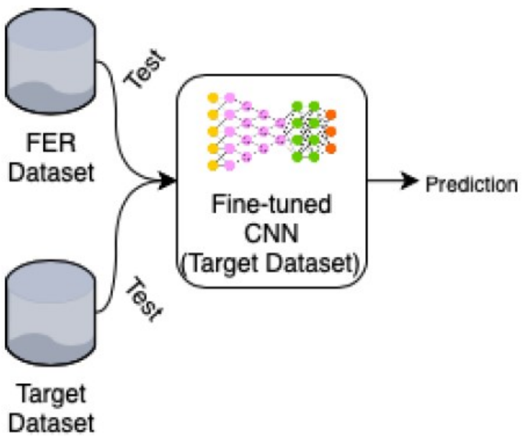
Method: Fine-tuning strategies

- Fig (a) - **FC strategy**: fine-tuning the full-connected layers of the CNN.
- Fig (b) - **FC+C strategy:** fine-tuning the full-connected layers + the last convolutional layer.
- Fig (c) - **RE strategy:** fine-tuning all layers by retraining the pre-trained model).



Experimental Protocol

- **Fine-tuned models are evaluated on both source and target datasets.**



Organization of the datasets

Dataset	Number of Samples and (%)			
	Total	Training	Validation	Test
FER	35,888	28,709 (80.0)	3,589 (10)	3,589 (10.0)
JAFFE	213	131 (61.5)	15 (7.0)	67 (31.5)
TFEID	510	405 (79.4)	47 (9.2)	58 (11.4)
MUG	7,352	5,877 (80.0)	741 (10)	734 (10.0)

- **Source:** FER dataset
- **Targets:** JAFFE, TFEID, MUG
- **Amount of training:** 50% and 100%
- **Training/validation/testing splitting:** a subject is present in only one subset.

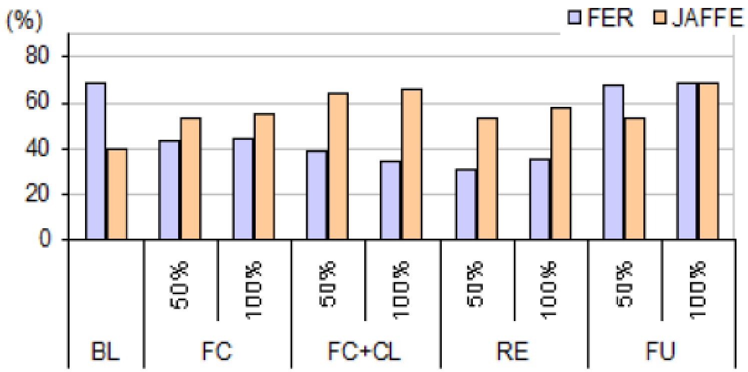
Experimental Results

- **Source: FER Target: JAFFE**
  - BL: Baseline model
  - FC: fine-tunning of the full-connected layers
  - FC+CL: fine-tunning of the full-connected layers + last convolutional layer
  - RE: retraining the pre-trained model
  - FU: fusion of both training datasets (source and target) for training the CNN.

	BL	FC		FC+CL		RE		FU	
		50%	100%	50%	100%	50%	100%	50%	100%
FER	68.35	43.08	44.61	39.09	34.22	30.76	35.36	67.69	**68.40**
JAFFE	40.30	53.73	55.22	64.18	65.67	53.73	58.21	53.73	**68.66**
FER	68.35	52.91	47.12	52.30	45.36	57.09	53.13	68.15	**68.77**
TFEID	31.03	67.24	82.76	60.34	**87.93**	56.90	86.03	63.79	86.21
FER	68.35	46.25	43.91	42.18	39.45	50.91	52.66	**68.99**	68.68
MUG	43.99	63.56	72.20	64.78	73.95	66.40	67.88	71.79	**81.24**

FC, FC+CL, RE: performance on the target data increases, while on the source data decreases when compared to BL.

FU: JAFFE has 219x less training samples than FER, the performance on the JAFFE test set boosted from 40.30% to 68.66%



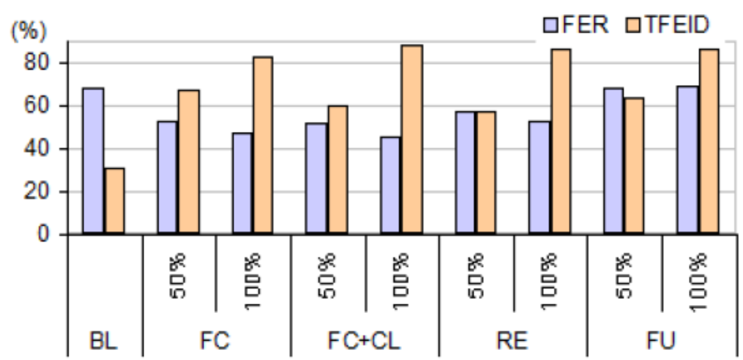
Experimental Results

- **Source: FER   Target: TFEID**
  - BL: Baseline model
  - FC: fine-tunning of the Full-Connected Layers
  - FC+CL: fine-tunning of the Full-Connected Layers + Last Conv. Layers
  - RE: model retraining from scratch
  - FU: fusion of training datasets (source and target)

	BL	FC		FC+CL		RE		FU	
		50%	100%	50%	100%	50%	100%	50%	100%
FER	68.35	43.08	44.61	39.09	34.22	30.76	35.36	67.69	**68.40**
JAFFE	40.30	53.73	55.22	64.18	65.67	53.73	58.21	53.73	**68.66**
FER	68.35	52.91	47.12	52.30	45.36	57.09	53.13	68.15	**68.77**
TFEID	31.03	67.24	82.76	60.34	**87.93**	56.90	86.03	63.79	86.21
FER	68.35	46.25	43.91	42.18	39.45	50.91	52.66	**68.99**	68.68
MUG	43.99	63.56	72.20	64.78	73.95	66.40	67.88	71.79	**81.24**

FC, FC+CL, RE: performance on the target data increases, while on the source data decreases when compared to BL.

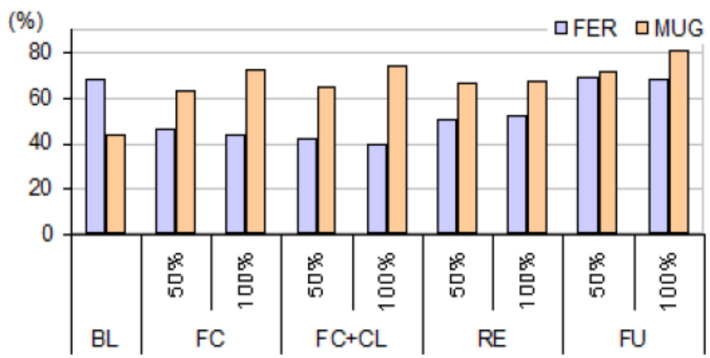
FU: TFEID has 71x less training samples than FER, the performance on the TFEID test set boosted from 31.03% to 86.21%



Experimental Results

- **Source: FER Target: MUG**
  - BL: Baseline model
  - FC: fine-tunning of the Full-Connected Layers
  - FC+CL: fine-tunning of the Full-Connected Layers + Last Conv. Layers
  - RE: model retraining from scratch
  - FU: fusion of training datasets (source and target)

	BL	FC		FC+CL		RE		FU	
		50%	100%	50%	100%	50%	100%	50%	100%
FER	68.35	43.08	44.61	39.09	34.22	30.76	35.36	67.69	**68.40**
JAFFE	40.30	53.73	55.22	64.18	65.67	53.73	58.21	53.73	**68.66**
FER	68.35	52.91	47.12	52.30	45.36	57.09	53.13	68.15	**68.77**
TFEID	31.03	67.24	82.76	60.34	**87.93**	56.90	86.03	63.79	86.21
FER	68.35	46.25	43.91	42.18	39.45	50.91	52.66	**68.99**	68.68
MUG	43.99	63.56	72.20	64.78	73.95	66.40	67.88	71.79	**81.24**



FC, FC+CL, RE: performance on the target data increases, while on the source data decreases when compared to BL.

FU: MUG has 5x less training samples than FER, the performance on the MUG test set boosted from 43.99% to 81.24%

Conclusions

- **We have investigated several approaches to supervised domain adaptation of CNNs for facial expression recognition.**
- **Answering our research questions**
  - What is the impact of different transfer learning strategies in the CNN memory integrity using cross-dataset?
    - In general, fine-tuning pre-trained CNNs does not preserve the memory integrity of the model.
    - The models adapted for the target dataset lose performance on the source dataset.
  - **Which action leads to the best performance on both source and target datasets?**
    - Fusion of source and target datasets has shown to be a promising strategy even when the target dataset is very small when compared with the source data.
  - **In fact, whether transfer learning is useful or not?**
    - Transfer learning is not always useful as we observed in the experiments. It is preferable to use a simpler architecture, more adapted to the problem than using a complex pre-trained network and transfer learning.

Thanks !

Any question?