Dream Net: a privacy preserving continual learning model for face emotion recognition

Marion Mainsant, Miguel Solinas, Marina Reyboz, Christelle Godin, Martial Mermillod

Dream Net: a privacy preserving continual learning model for face emotion recognition

Marion MAINSANT
CEA, LIST
Univ. Grenoble Alpes
F-38000 Grenoble, FRANCE
marion.mainsant2@cea.fr

Miguel SOLINAS
CEA, LIST
Univ. Grenoble Alpes
F-38000 Grenoble, FRANCE
miguelangel.solinas@cea.fr

Marina REYBOZ
CEA, LIST
Univ. Grenoble Alpes
F-38000 Grenoble, FRANCE
marina.reyboz@cea.fr

Christelle GODIN
CEA, LETI
Univ. Grenoble Alpes
F-38000 Grenoble, FRANCE
christelle.godin@cea.fr

Martial MERMILLOD
LPNC
Univ. Grenoble Alpes & CNRS
F-38000 Grenoble, FRANCE
martial.mermillod@univ-grenoble-alpes.fr

Abstract— **Continual learning is a growing challenge of artificial intelligence. Among algorithms alleviating catastrophic forgetting that have been developed in the past years, only few studies were focused on face emotion recognition. In parallel, the field of emotion recognition raised the ethical issue of privacy preserving. This paper presents Dream Net, a privacy preserving continual learning model for face emotion recognition. Using a pseudo-rehearsal approach, this model alleviates catastrophic forgetting by capturing the mapping function of a trained network without storing examples of the learned knowledge. We evaluated Dream Net on the Fer-2013 database and obtained an average accuracy of 45% ± 2 at the end of incremental learning of all classes compare to 16% ± 0 without any continual learning model.**

***Index Terms*— continual learning, incremental learning, pseudo-rehearsal, catastrophic forgetting, privacy, face emotion recognition, replay method**

I. INTRODUCTION

Emotion plays a central role in many social interactions. During their exchanges, human beings use the tone of their voice, facial expressions or even gestures to convey their feelings and use these same universal keys to decode the emotions of their peers. Computers, as "smart" as they may be, do not yet have access to this essential capability of human communication. As facial expressions are one of the main cues for human non-verbal communication, emotion recognition through facial images have been widely studied in the last decades [1], [2]. Deep learning methods are increasingly used for face emotion recognition in the wild because they show very high performances extracting and analyzing features from raw data [3].

The correct training of deep learning algorithm requires many input data and such databases are not easy to design. Laboratory databases mainly cover the seven basis emotions that are Neutral, Anger, Fear, Sadness, Disgust, Surprise and Happiness. But other finer emotion also exists [4], [5] and new methods like FACS [6] were developed in order to cover this larger range of emotional expression. Most of facial emotion recognition systems use Artificial Neural Networks (ANNs) trained "offline" with such databases. However, to move towards more and more intelligent systems adapting to their environment, we are going to encounter the problem of continual learning with new examples or with new emotions.

From human point of view, this continuous adaptation to new information is possible as people continually learn from their own experiences with the ability to keep and fine-tune previously acquired knowledge. In ANNs, with such constraint, continual adaptation is not possible by default. When a trained ANN learns new examples, it adapts its parameters in order to fit the new set of examples without taking into consideration previous knowledge. Consequently, the ANN no longer fits the previous examples, leading to a drastic reduction of the model's performance. That effect is called "catastrophic forgetting" and is a major issue in deep learning [7], [8]. In ANN continual learning, the easiest way to overcome catastrophic forgetting is to learn new training examples jointly with old ones to avoid forgetting previously seen examples. In this way, the best and simplest solution is to store all the previously seen examples. However, this solution is unrealistic for three main reasons:

i. Privacy issues are usually a concern when storing raw proprietary data. In this paper, we want to propose a way to overcome this important ethical problem.
ii. Large memory footprint requirements are often impractical for edge or embedded devices.
iii. Complete retraining for each new set of incoming data is infeasible on large scales due to computational power and learning time limitation.

Since the 90s, various algorithms have been developed in order to deal with continual learning issues, we are now able to separate the different approaches into three groups [9]: regularization inspired from synaptic consolidation, parameter-isolation inspired from neurogenesis and replay inspired from human memory consolidation. Regularization main characteristic is to distribute the knowledge over the network by maintaining most important weights of the previously trained network while learning new classes [10], [11]. Parameter isolation method consist in freezing the ANN parameters and allocating additional neural resources to

acquire new knowledge [12], [13]. Replay methods mainly consists in rehearsing old knowledge when learning a new task [14], [15].

One of the difficulty of continual learning approaches is to find a trade-off between stability and plasticity, the dilemma at the origin of catastrophic forgetting issues. In fact, in an ANN, plasticity of connections is essential for the learning of new knowledge while it needs stability in order to conserve the previous encoding knowledge [16], [17]. Regularization and parameter isolation approaches most of the time trade their plasticity for stability. We will thus focus on replay methods that nowadays find a better answer to this stability-plasticity dilemma. Replay methods are inspired from the human brain hippocampal-neocortical system. In fact, some studies suggest that one of the main source of memory consolidation in human brain is the replay of neural activity examples during sleep [8], [18]. It exists two ways to implement replay methods: the rehearsal approach that stores a fraction of old examples [14], [19] and the pseudo-rehearsal approach that uses artificially generated examples representing previously learned knowledge [15], [20]. In both cases, those examples are then interlaced with new ones during the learning of new examples.

A replay model combining rehearsal and pseudo-rehearsal approaches has been recently proposed [21]. In this model, examples representing previously learned knowledge are generated through a reinjection sampling procedure (i.e. iterative sampling [22]) that uses old examples stored in a small buffer as a seed.

All those strategies implicitly assume that examples that resemble the input distribution are necessary for optimal retrieval of old knowledge. However, only a few works have focused on an optimal model for continual learning with privacy constraints.

In the field of emotion recognition, privacy is a major ethical issue since the face is one of the more useful biometric data to recognize a person on images [23]. For this reason, this study proposes a privacy-preserving continual learning model for facial emotion recognition, Dream Net. This model is a data free version of the combined-replay model [21]. To preserve the privacy of the data, we improved the model to employ the same reinjection sampling procedure but using random noise as seed.

In the next section (II) we provide more background on continual learning and emotion detection. Then we describe the Dream Net model in section III. In section IVwe present results of our experiments. Finally, section V is about conclusion and perspective.

II. Related work

A. Catastrophic forgetting and continual learning

Catastrophic forgetting is a challenging issue of deep learning related to the stability-plasticity dilemma of ANNs [16]. As explained in the introduction (I), the most effective approaches to avoid catastrophic forgetting in continual learning problems are based on replay strategies that find an optimal compromise between stability and plasticity [24]. For example, a traditional way to alleviate catastrophic forgetting is to relearn a minimal portion of what the model has learned before. This is the case of the Episodic replay algorithm proposed by Chaudhry *et al.* [14]. This approach uses a single ANN and a tiny memory buffer that stores old examples and associated labels. For each new example, the ANN is trained with old examples from the memory buffer alongside the new examples.

Alternatively, the replay strategy, introduced by Robins et al. [25], consists of replaying what the neural network might have learned before by employing an input stimulus and the associated activation pattern at the output of the network instead of the ground-truth labels. Various study highlighted that this input-output activation patterns contain enough information to prevent the ANN from catastrophically forgetting previously acquired knowledge [19], [26]. Replay methods are usually divided into two categories: rehearsal and pseudo-rehearsal. In both strategies, the idea is to capture and replay the learned predictive function *f*, which is the function encoded by the ANN when it learns to associate a set of input examples to its corresponding outputs labels [20], [25]–[27].

Rehearsal methods aim to capture *f* through real examples of a memory buffer. At each learning step, a portion of learned examples is stored in a memory buffer in order to be used later in the process of capturing *f*. Icarl [19] was the first model to use this solution in large-scale datasets. This model uses a memory buffer and a knowledge distillation strategy [28] to retrieve and consolidate previously learned knowledge. It is among the best continual learning solution in the state of the art with Episodic replay algorithm [14].

Instead of using real-examples from previously learned knowledge, the pseudo-rehearsal method consists in generating with a second network, an auxiliary set of examples that represents the original input distribution. In this way, the generated synthetic examples and their corresponding activation patterns are employed to consolidate previously acquired knowledge [15], [26], [29], [30]. This approach has been improved by the development of powerful generative models such as Generative Adversarial Networks (GAN) [31] and Variational Auto-Encoder (VAE) [32]. However, recent works raise the issue of the instability of generative models, the challenges in modeling complex distributions [33] and the potential privacy issues [24]. Alternatively to those generative models, Rousset *et al.* [26] proposed a pseudo-rehearsal method that bypasses the input distribution by employing a sampling procedure to capture the learned function without relying on realistic examples (i.e. examples that look like the previously learned samples). Their solution relies on a dual network architecture (Net1 and and Net2). Net1 is in contact with a new input stimuli, while Net2 generates synthetic examples for Net1, allowing Net1 to preserve old knowledge. Synthetic examples generation in Net2 is done with a re-injection sampling procedure (i.e. iterative sampling) that enables to capture the learned function. The interest of this method is to capture the network's learning function with synthetic examples instead of representing previously seen examples.

An approach combining advantages of rehearsal and pseudo-rehearsal approach was developed recently: the combined replay model [21]. This model uses the tiny memory buffer with real examples to capture previously learned knowledge through the reinjection sampling procedure proposed in [22], [29]. This approach shows

competitive results for tiny memory buffers on Mnist, Cifar10 and Cifar100 that are classic deep learning datasets.

Excepted for Rousset *et al.* [26] most of the solutions presented above implicitly assume that examples that look like the input distribution are needed to capture *f*.
Inspired by Rousset *et al.* [26] and by the combined replay approach [21], our work focuses on building a model with an optimal architecture to capture the mapping function without relying on input distribution. We thus propose a totally data-free model, Dream Net, that alleviates catastrophic forgetting without relying on small buffers or generative models.

B. *Continual learning for emotion recognition*

To the best of our knowledge, only two recent articles propose continual learning algorithms for emotion recognition that alleviate catastrophic forgetting.
The first one is an unsupervised setting for continual learning of emotions of individual persons on video and audio data [34]. To deal with catastrophic forgetting, the authors designed what they call an "affective memory" with a "growing when required" (GWR) network that is part of parameter-isolation continual learning approach. The proposed algorithm is very useful in the frame of emotion detection since it allows dealing with temporal audio and video data while providing emotion recognition. Nevertheless, the employed algorithm is part of the parameter-isolation continual learning approach which does not provide good trade-off between stability and plasticity.
The second one presents a dual-memory framework for continual learning of facial emotion [35]. They designed an auto-encoder based "imagination" model for continual learning of emotions on facial expression. The goal of this framework is to simulate the human capability to imagine interaction in order to improve its abilities to remember previously seen expressions and generalizing on unseen ones for a given subject. The generative part of the model allows to create photo-realistic images of the six basis emotion expression for a given face image. It allows the creation of additional data for each class in order to increase feature representation and thus consolidate the learning process. The dual memory framework is composed of two "growing when required" networks representing respectively semantic and episodic memory. Episodic memory sequentially receives information and creates a feature prototype from input. Then, the semantic memory receives "winner neurons" from the episodic memory and add them to its architecture only if they are more accurate than existing ones. A pseudo-rehearsal process enables to replay periodically episodic activations in order not to forget representations of previous classes. This continual learning framework is not easy to deploy because it uses both neurogenesis and pseudo-rehearsal approaches. With our proposed model Dream Net, the approach is different; we do not train our model for a specific subject but on a generic database. The strengths of our approach compared to existing ones in emotion recognition field are that Dream Net is an agnostic model independent of the database and privacy preserving by design.

III. METHOD

This section presents the continual learning Dream Net designed to overcome catastrophic forgetting in a privacy preserving way. This section begins with the database choice and associated feature extraction method. We then detail Dream Net architecture and finally present the experiments carried out.

A. *Database*

To assess our model in the field of face emotion recognition, we worked with the Fer-2013 database [36]. It contains 35685 grayscale 48x48 pixels images with all the basis emotions and covering all age, gender and ethnicity (Figure 1). This dataset is particularly useful for deep-learning as it is a good compromise between a large database like AffectNet [37] which would add a scaling problem and smaller ones like CK+ [38] or Jaffe [39] captured in a laboratory environment that provide only posed emotions. Furthermore, Standford University published an open source study which aims to calibrate a ResNet50 feature extractor to Fer-2013 database [40] inspired from Pramerdorfer et al. work [41]. Choosing Fer-2013 dataset thus enabled us to begin our study with a baseline architecture that corresponds to the best of the state of the art for Fer-2013.

Figure 1: Examples of Fer-2013 database [36]

TABLE 1 shows Fer-2013 images distribution over class. The emotion "disgust" is under-represented while the emotion "happy" is over-represented compare to other emotions in the dataset. We firstly split the dataset into training (80% of the database), validation (10% of the database) and test (10% of the database) sets so as not to test the model with training data. Then, in order to obtain a balanced training dataset that contains 5000 images per classes, we randomly duplicate images of under-represented classes and delete images in over-represented classes. This method prevents the introduction of a bias due to the imbalance of classes during the training step, its interest has been shown in the two following papers [42], [43].

TABLE 1: FER-2013 EMOTION DISTRIBUTION OVER TRAINING, VALIDATION AND TEST SETS (80%, 10%; 10%)

Emotion						
Angry	***Disgust***	***Fear***	***Happy***	***Sad***	***Surprise***	***Neutral***
4955	549	5124	8992	6080	4005	6201

B. *Feature extraction*

For feature extraction, we employ the ResNet50 model pre-trained on the Fer-2013 database delivered by Stanford University study [40]. Figure 2 details the ResNet50 architecture, the input of the network takes Fer-2013 images resized to 197x197 pixels on RGB channels and then a

succession of convolutional and identity blocks allows the extraction of features in a way to classify facial emotions. The network is provided with the classifier that allowed its pre-training. In order to use the ResNet50 network only for feature extraction, we cut it at the red mark on Figure 2, just before the classifier input. For each image of the Fer-2013 database, we thus obtain 2048 features that represent the facial emotion of the person. Those features are then incrementally given to the input of the Dream Net model.

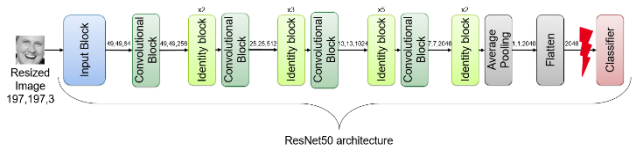

Figure 2: Pre-trained ResNet50 architecture
Convolutional Block extracts features while changing the input dimension.
Identity Block enables to extract features without changing input dimension.

In this work, to ensure that the convolutional filters are well trained, we employ the presented ResNet-50 pre-trained on Fer-2013. Our choice may be arguable since we employ the most optimal convolutional filters; however, feature representation is beyond this paper's scope.

C. *Dream Net architecture*

Dream Net is a data-free version of the combined replay continual learning algorithm proposed in [21]. The original architecture is inspired from Rousset and Ans work [26] which consists in alleviating catastrophic forgetting using a pseudo-rehearsal incremental technique that captures the model's learned function. In the following part, the term *pseudo-example* refers to a pair (pseudo-feature, pseudo-label) artificially generated that characterizes the learned function of a model.

Figure 3 illustrates the global architecture of the model composed of two fully connected hybrid networks: *Learning Net* and *Memory Net* which are structured as following:

- An input layer that has the size of features extracted from images.
- Several hidden layers with parameters depending on the considered database. For Fer-2013 we use one hidden layer with 1000 neurons.
- An output layer, with sigmoid activation function, composed of several neurons corresponding to the input (Auto-associative or Auto-encoder part) and several neurons corresponding to the number of classes (Hetero-associative or part).

The hybrid architecture is also called Auto-Hetero associative ANN because it allows replicating input information like a standard auto-encoder (*Auto*) and classifying data in a supervised way like a standard classifier (*Hetero*) in a single inference. This particular architecture enables the algorithm to learn the dataset and generate pseudo-examples that capture the learned function.

Our Dream Net architecture can be divided into three phases (mentioned as 1, 2 and 3 on Figure 3):

1) *Learning Net* learns real-features from a class N and pseudo-features from the previously learned classes (0 to N-1)

2) *Learning Net* transfers its weights to *Memory Net*.

3) *Memory Net* captures the learned function using a reinjection sampling procedure. The reinjection sampling procedure consists in the following steps: inject a random noise input vector and reinject the replication vector obtained at the output of the auto-associative part of *Memory Net* at its input and so on. At each reinjection, Auto and Hetero associative outputs of *Memory Net* are conserved to create pseudo-examples. After several reinjection, we obtain a pseudo-examples database that contains pseudo-features and corresponding pseudo-labels obtained after each reinjection (data from the first inference is not kept). We note that for the first class learned, an un-trained *Memory Net* also generates pseudo-examples.

For each new class to learn, we repeat this learning cycle.

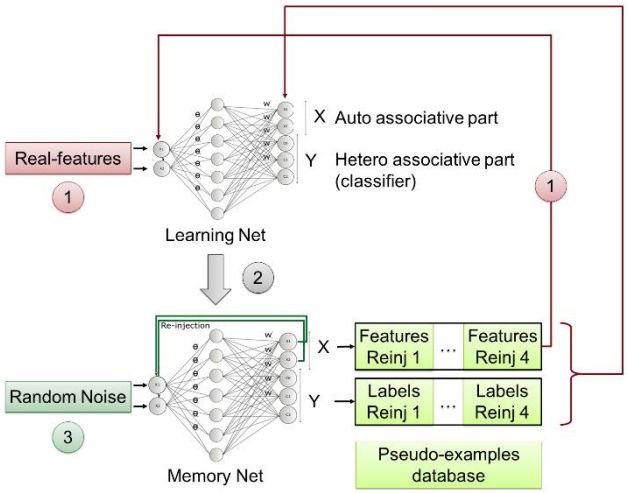

Figure 3: Dream Net model architecture scheme

D. *Experiments*

To benchmark our model in front of state of the art incremental learning methods, we implemented several models with the Fer-2013 database. We separate those algorithms into two categories: baseline models and literature models.

Baseline models:

Baseline models are models not implemented in literature but that enable to position our accuracy results and show that effects observed are due to our model architecture specificities and not to other properties. These are therefore control models.

- Offline:

This first architecture consists in having an auto-hetero associative ANN like *Learning Net* and *Memory Net* and train it with all classes at the same time. We designed this network such as we obtain the same accuracy results on Fer-2013 database as with the simple classifier initially implemented in the ResNet50 network used for feature extraction (Figure 2). This enables to show that the hybrid architecture does not have a detrimental effect on the final accuracy and gives us the maximum accuracy we can obtain at each step of the learning.

- Class by class learning without specific algorithm:

The goal of this second architecture is to highlight the catastrophic forgetting effect. For this, we use a single Auto-Hetero associative ANN without reinjection sampling procedure. The only action to remember previously learned classes is to initialize the network with previously learnt weights. The model receives the classes to learn one by one.

- Auto-Hetero replay:

For the third architecture, we use two Auto-Hetero associative ANN with similar role as *Learning Net* and *Memory Net* in the Dream Net model but without the reinjection sampling procedure on *Memory Net*. This last baseline model enables to highlight the relevance of reinjection with respect to dual-network architecture for the success of the algorithm.

Literature models:

Literature models are state of the art incremental learning algorithms that have currently the best results on classic machine learning databases: Mnist, Cifar10 and Cifar100.

- Icarl [19]:

We implemented a fully connected version of Icarl [15] which is a rehearsal method that also uses a dual-network. Networks used in this algorithm are Hetero associative ANN (classifiers). One of them captures previously learned knowledge using a memory buffer composed of old real examples. This process is called classifier-based distillation. We decided to compare our model to this one due to its superior performances compare to other incremental learning algorithms [9].

- Episodic replay [14]:

This algorithm uses a simple classifier with a memory buffer composed of old real examples in order to remember previously learned classes. We implemented the fully connected version of this method that is currently listed in incremental learning state of the art as better than other replay methods.

- Combined replay [21]:

Dream Net is the data free version of this model. Consequently, its architecture is very close to Dream Net presented above in subsection III-C. The major difference is that a tiny memory buffer composed of old real examples is used for the reinjection sampling procedure instead of random noise.

Hyper-parameters and Metrics:

We perform all the experiments with the hyper-parameters presented in TABLE 2 for all presented algorithms.

The random noise use for our model, Dream Net, is an isotopic Gaussian distribution with a center at 0 and a variance of 1, N(0, 1). We chose this distribution because this is the one classically used in generative models [32], [44].

We measure the performance of all our experiments on the testing set using accuracy. All our results are averaged over 10 runs and we display confidence intervals at 95% on all of them.

TABLE 2: MODELS HYPER-PARAMETERS

Models	**Hyper parameters**			
	Units per hidden layer	***Activation function***	***Optimizer***	***Last layer activation***
Offline	[4096, 1024]	[relu, relu]	Adam	Sigmoid
Auto-Hetero Replay	[1000]	[relu]	Adam	Sigmoid
Icarl	[1000, 1000]	[Mish, Mish]	SGD*	Sigmoid
Episodic Replay	[1000, 1000]	[Mish, Mish]	SGD*	Softmax
Dream Net model	[1000]	[relu]	Adam	Sigmoid

*SGD = Stochastic Gradient Descent

IV. RESULTS

In this section, we present all experimental results obtained. First, we evaluate Dream Net accuracy on several relevant parameters. Then, we study the influence of emotion order and we finally present a general benchmark that enables to position the Dream Net model in relation to literature and baseline models.

For the study of Dream Net parameters and for the general benchmark, emotion order is arbitrarily fixed to [angry, disgust, fear, happy, sad, surprise, neutral].

A. *Dream Net parameters tuning*

- Ratio pseudo-examples over real-examples effect:

This first study enables us to choose the ratio of pseudo-examples over real-examples. In fact, as the learning of the last emotion consists in interlacing real examples of this emotion with pseudo-examples representing the learned function of previous emotions, we wondered if increasing the number of pseudo-examples compared to real-examples could improve performances on previously learned emotions. We thus evaluated the impact of the ratio modification on Dream Net accuracy in Figure 4. To change the ratio we increased or decreased the number of generated pseudo-examples by changing the number of random noise batches at the input of *Memory Net*.

Figure 4 shows that from a ratio of 20 (i.e. 1 true example for 20 pseudo-examples), the increase of the ratio does not increase the global accuracy of Dream Net. We can even observe a decrease of the global accuracy from a ratio of 40. This curve thus highlights that the ratio of 20 is optimal and increasing it does not improve the accuracy of the model.

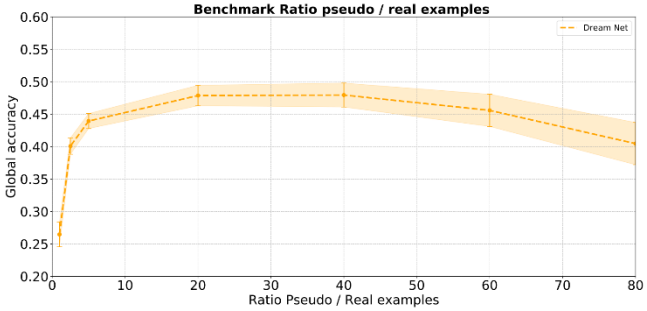

Figure 4: Ratio pseudo vs real examples study.
Final global accuracy of Dream Net over pseudo vs real examples ratio.

- Number of reinjections effect:

We then studied the effect of the number of reinjection in the reinjection sampling procedure of *Memory Net*. We tested six different numbers of reinjection between 0 and 10.

Figure 5 gives the global accuracy of the model after continual learning of the seven emotions over the number of reinjections. We can see that we reach the optimal accuracy for Dream Net for four reinjections. From this value, the increase of the number of reinjections does not changes the accuracy of Dream Net model. We can thus conclude that four reinjections are optimal for Dream Net.

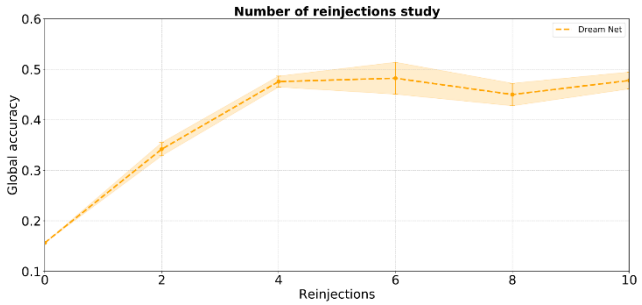

Figure 5: Reinjection number study.
Final global accuracy of Dream Net over the number of reinjection.

B. *Influence of emotion order*

We then studied the emotion order influence on Dream Net model accuracy. The last emotion was fixed and another six emotions order was randomly chosen at the beginning of each run. Figure 6 summarizes obtained results and gives each emotion average performance depending on last learned emotion. It clearly appears that emotions perceptually easier to recognize (i.e. happy and surprise emotions) are generally better classified than other emotions [45], [46]. In fact, we observe that emotions that had already been recognized better when learned offline are also better remembered during continual learning. Besides, we notice that the last emotion learned is always better memorized than the others. Nevertheless, the order of emotion does not have a significant impact on the global accuracy of the model at the end of the learning. Dream Net thus increases emotion accuracy disparities already observed when training offline and favors the last learned class.

		Emotion average performance							Global accuracy
		Angry	Disgust	Fear	Sad	Surprise	Happy	Neutral	
Last emotion	Angry	91% ± 1	27% ± 6	17% ± 11	12% ± 8	60% ± 14	68% ± 5	27% ± 14	43% ± 3
	Disgust	29% ± 15	93% ± 2	22% ± 15	21% ± 14	58% ± 14	66% ± 7	56% ± 21	49% ± 2
	Fear	23% ± 9	29% ± 7	89% ± 2	5% ± 3	35% ± 5	65% ± 5	27% ± 10	39% ± 2
	Sad	26% ± 3	35% ± 3	17% ± 2	92% ± 0	50% ± 4	67% ± 3	19% ± 2	44% ± 1
	Surprise	38% ± 6	56% ± 5	7% ± 2	34% ± 7	92% ± 0	64% ± 2	40% ± 6	47% ± 1
	Happy	28% ± 5	49% ± 4	13% ± 5	39% ± 6	63% ± 4	96% ± 0	32% ± 6	46% ± 2
	Neutral	37% ± 5	36% ± 5	25% ± 4	12% ± 3	46% ± 5	61% ± 2	95% ± 0	45% ± 2
Mean for each emotion		39% ± 1	46% ± 1	27% ± 2	31% ± 2	58% ± 1	69% ± 1	42% ± 2	45% ± 1
Accuracy Offline		61%	75%	51%	60%	81%	87%	71%	69%

Figure 6: Dream Net emotion order influence.
For each fixed last emotion, 10 runs with random different order for the six first learned emotions have been done.

C. *General benchmark*

Figure 7 summarizes the results of our general benchmark that compares Dream Net performances with baseline and literature models. The light blue curve of "Class by class without specific algorithm" shows the catastrophic forgetting effect discussed in section II. In this algorithm, a single auto-hetero associative ANN is trained class by class and we can observe that global accuracy falls until 16%. We fixed Icarl, Episodic replay and Combined replay memory buffer sizes to 10 real-examples per class which gives a total size of 70. We can see on Figure 7 that Dream Net overcomes catastrophic forgetting and is significantly above Icarl and Episodic replay for this memory buffer size. As "Auto-Hetero noise replay" model is superimposed with "Class by class without specific algorithm" model, we can conclude that the model results are not due to dual-network architecture. Thus, the reinjection sampling procedure is consequently at the origin of Dream Net performances. Besides, we can see that combined replay accuracy is not significantly above Dream Net. Consequently, using random noise for the reinjection sampling procedure is similar as using a memory buffer of 70 real examples stored from previously learned classes.

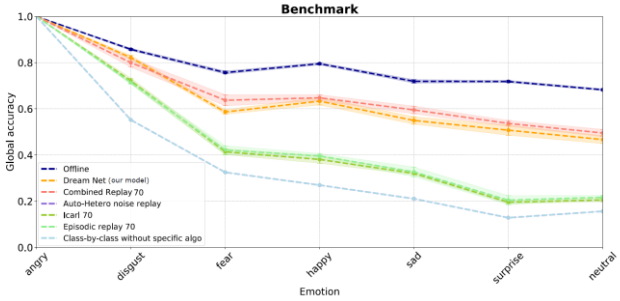

Figure 7: General accuracy benchmark with baseline and literature models.
Y-axis represents the global accuracy of the model on all learned classes.
X-axis gives the last added emotion.

In order to compare more accurately Dream Net performance to Icarl, Episodic replay and Combined replay performances, Figure 8 shows the global final accuracy of each model over memory buffer sizes. As explained in III, Dream Net does not require a memory buffer, that is why its accuracy is constant. We can notice that until a memory size of 1400 real examples Icarl and Episodic replay are below or not significantly above Dream Net. A memory size of 1400 real-examples stored is not negligible as it represents a third of training examples of a class. Dream Net is thus a compromise to obtain good performances without storing any real-example from the database, which means that our model is agnostic in the sense that it can learn without a priori knowledge on the data and offers a privacy-preserving solution.

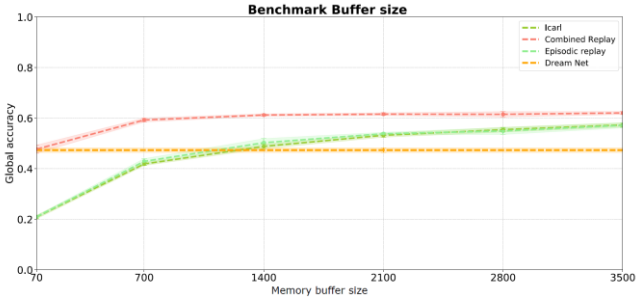

Figure 8: Memory buffer size benchmark.
Accuracy of Combined replay, Icarl and Episodic replay models over memory buffer size after learning all classes incrementally.

In summary, Dream Net outperforms literature models for memory buffer smaller than 1400 real example stored. Its final accuracy is significantly above the one associated with the original catastrophic forgetting effect. This model thus appears to be a robust privacy preserving alternative to replay models that store old examples in memory buffers.

V. Conclusion

This paper presents Dream Net, the first privacy preserving continual learning model for face emotion recognition. This model is an extension of the already existing continual learning method, combined replay, designed to overcome catastrophic forgetting which has the particularity of not storing any previously learned data, and the specificity to be agnostic concerning the class to learn. Moreover, this method, unlike other continual learning methods, does not require the creation of new synaptic weights, new neurons or new multi-head networks. Experimental results on Fer-2013 database lead to the following conclusions.

The proposed Dream Net model overcomes catastrophic forgetting when learning incrementally new classes and outperforms Icarl and Episodic replay literature models for memory buffer of size bellow 1400. We highlighted in section IV that this performance is due to the reinjection sampling procedure used to create the pseudo-examples database.

Even if Dream Net does not overcome literature models for all memory size and even if the combined replay model gives more accurate results for a buffer size over 70, this model has the great advantage to be totally data free. This property could be of crucial importance for real applications involving privacy preserving issues. For instance, if we want to learn a new emotion with a network already trained on other emotions, our method will protect the privacy of people who participated to the development of the first database of emotion. In the present paper we only present a study of class-by-class continual learning because it is a common approach to benchmark these types of model. In future work we plan to investigate other type of scenario, closer to "real-life" problematics:

- Learn a database containing the seven basis emotions and extend incrementally the model to finer emotions without forgetting basis emotions.
- Learn new examples of classes already present and therefore strengthen the classifier with examples learned over time without retaining the examples already learned. With this kind of streaming scenario, overcoming the privacy issue thanks to Dream Net will take on its full meaning.

In the emotion order study (Figure 6) we shown that, even if the emotion order does not have a drastic influence on the accuracy of each emotion nor on the global accuracy, last emotion learned is significantly better memorized than emotions learned before. As we already fine-tuned the ratio between pseudo-examples of previously learned classes and real-examples of the new class and the number of reinjections in the reinjection sampling procedure, those parameters do not allow improving results' accuracy. A promising possibility for future experiment will be to test more optimal seeds. We thus plan to investigate more precisely the choice of initial noise at the input of *memory net* in order to improve performances.

Future work will also include the adaptation of Dream Net to more complex emotional databases in order to test the generalizability of our model.

Acknowledgment

This work has been partially supported by MIAI@Grenoble Alpes, (ANR-19-P3IA-0003)

References

[1] J. Kumari, R. Rajesh, and K. M. Pooja, "Facial Expression Recognition: A Survey," *Procedia Comput. Sci.*, vol. 58, pp. 486–491, 2015

[2] Dhwani Mehta, Mohammad Siddiqui, and Ahmad Javaid, "Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality," *Sensors*, vol. 18, no. 2, Art. no. 2, Feb. 2018

[3] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Trans. Affect. Comput.*, pp. 1–1, 2020

[4] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, Art. no. 1, Jan. 2009

[5] R. S. Deshmukh and V. Jagtap, "A survey: Software API and database for emotion recognition," in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, Jun. 2017, pp. 284–289.

[6] F. De la Torre and J. F. Cohn, "Facial Expression Analysis," in *Visual Analysis of Humans: Looking at People*, T. B. Moeslund, A. Hilton, V. Krüger, and L. Sigal, Eds. London: Springer, 2011, pp. 377–409.

[7] M. McCloskey and N. J. Cohen, "Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem," in *Psychology of Learning and Motivation*, vol. 24, G. H. Bower, Ed. Academic Press, 1989, pp. 109–165.

[8] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory.," *Psychol. Rev.*, vol. 102, no. 3, Art. no. 3, Jul. 1995

[9] M. De Lange *et al.*, "A continual learning survey: Defying forgetting in classification tasks," May 2020

[10] F. Zenke, B. Poole, and S. Ganguli, "Continual Learning Through Synaptic Intelligence," Jun. 2017

[11] J. Kirkpatrick *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proc. Natl. Acad. Sci.*, vol. 114, no. 13, Art. no. 13, Mar. 2017

[12] A. A. Rusu *et al.*, "Progressive Neural Networks," Sep. 2016

[13] G. Hocquet, O. Bichler, and D. Querlioz, "OvA-INN: Continual Learning with Invertible Neural Networks," Jun. 2020

[14] A. Chaudhry *et al.*, "On Tiny Episodic Memories in Continual Learning," Jun. 2019

[15] R. Kemker and C. Kanan, "FearNet: Brain-Inspired Model for Incremental Learning," Feb. 2018

[16] W. C. Abraham and A. Robins, “Memory retention – the synaptic stability versus plasticity dilemma,” *Trends Neurosci.*, vol. 28, no. 2, Art. no. 2, Feb. 2005

[17] M. Mermillod, A. Bugaiska, and P. Bonin, “The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects,” *Front. Psychol.*, vol. 4, 2013

[18] R. C. O’Reilly, R. Bhattacharyya, M. D. Howard, and N. Ketz, “Complementary Learning Systems,” *Cogn. Sci.*, vol. 38, no. 6, pp. 1229–1248, 2014

[19] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “iCaRL: Incremental Classifier and Representation Learning,” Apr. 2017

[20] T. Lesort, H. Caselles-Dupré, M. Garcia-Ortiz, A. Stoian, and D. Filliat, “Generative Models from the perspective of Continual Learning,” Dec. 2018

[21] M. Solinas *et al.*, “Beneficial Effect of Combined Replay for Continual Learning:,” in *Proceedings of the 13th International Conference on Agents and Artificial Intelligence*, Online Streaming, --- Select a Country ---, 2021, pp. 205–217.

[22] M. Solinas, C. Galiez, R. Cohendet, S. Rousset, M. Reyboz, and M. Mermillod, “Generalization of iterative sampling in autoencoders,” p. 9.

[23] A. Agarwal, P. Chattopadhyay, and L. Wang, “Privacy preservation through facial de-identification with simultaneous emotion preservation,” *Signal Image Video Process.*, Nov. 2020

[24] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer, “Class-incremental learning: survey and performance evaluation,” Oct. 2020

[25] A. Robins, “Catastrophic Forgetting, Rehearsal and Pseudorehearsal,” *Connect. Sci.*, vol. 7, no. 2, Art. no. 2, Jun. 1995

[26] B. Ans and S. Rousset, “Avoiding catastrophic forgetting by coupling two reverberating neural networks,” *Comptes Rendus Académie Sci. - Ser. III - Sci. Vie*, vol. 320, no. 12, Art. no. 12, Dec. 1997

[27] F. Lavda, J. Ramapuram, M. Gregorova, and A. Kalousis, “Continual Classification Learning Using Generative Models,” Oct. 2018

[28] G. Hinton, O. Vinyals, and J. Dean, “Distilling the Knowledge in a Neural Network,” Mar. 2015

[29] G. M. van de Ven, H. T. Siegelmann, and A. S. Tolias, “Brain-inspired replay for continual learning with artificial neural networks,” *Nat. Commun.*, vol. 11, no. 1, p. 4069, Dec. 2020

[30] R. M. French, “Pseudo-recurrent Connectionist Networks: An Approach to the ‘Sensitivity-Stability’ Dilemma,” *Connect. Sci.*, vol. 9, no. 4, Art. no. 4, Dec. 1997

[31] M. Zhai, L. Chen, F. Tung, J. He, M. Nawhal, and G. Mori, “Lifelong GAN: Continual Learning for Conditional Image Generation,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), Oct. 2019, pp. 2759–2768.

[32] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” May 2014

[33] T. Lesort, “Apprentissage continu: S’attaquer à l’oubli foudroyant des réseaux de neurones profonds grâce aux méthodes à rejeu de données,” p. 174.

[34] P. Barros, G. I. Parisi, and S. Wermter, “A Personalized Affective Memory Neural Model for Improving Emotion Recognition,” May 2020

[35] N. Churamani and H. Gunes, “CLIFER: Continual Learning with Imagination for Facial Expression Recognition,” p. 7.

[36] I. J. Goodfellow *et al.*, “Challenges in representation learning: A report on three machine learning contests,” *Neural Netw.*, vol. 64, pp. 59–63, Apr. 2015

[37] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, Art. no. 1, Jan. 2019

[38] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, San Francisco, CA, USA, Jun. 2010, pp. 94–101.

[39] I. M. Revina and W. R. S. Emmanuel, “A Survey on Human Face Expression Recognition Techniques,” *J. King Saud Univ. - Comput. Inf. Sci.*, p. S1319157818303379, Sep. 2018

[40] A. Khanzada, C. Bai, and F. T. Celepcikay, “Facial Expression Recognition with Deep Learning,” p. 6.

[41] C. Pramerdorfer and M. Kampel, “Facial Expression Recognition using Convolutional Neural Networks: State of the Art,” p. 7.

[42] T. T. D. Pham and C. S. Won, “Facial Action Units for Training Convolutional Neural Networks,” *IEEE Access*, vol. 7, pp. 77816–77824, 2019

[43] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018

[44] H. Shin, J. K. Lee, J. Kim, and J. Kim, “Continual Learning with Deep Generative Replay,” Dec. 2017

[45] M. N. Dailey, G. W. Cottrell, C. Padgett, and R. Adolphs, “EMPATH: A Neural Network that Categorizes Facial Expressions,” *J. Cogn. Neurosci.*, vol. 14, no. 8, pp. 1158–1173, Nov. 2002

[46] M. Mermillod, P. Bonin, L. Mondillon, D. Alleysson, and N. Vermeulen, “Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation,” *Neurocomputing*, vol. 73, no. 13–15, pp. 2522–2531, Aug. 2010