UNIVERSITY OF CAMBRIDGE
Department of Computer Science and Technology

15th IEEE International Conference on Automatic Face and Gesture Recognition

CLIFER: Continual Learning with Imagination for Facial Expression Recognition


Nikhil Churamani Hatice Gunes

UKRI Engineering and Physical Sciences Research Council

Motivation

Generalisation for Facial Expression Recognition (FER) *in-the-wild*

(a)

trade-off

Personalisation to Learn Individual Expressions

(b)

(a) https://ibug.doc.ic.ac.uk/resources/first-affect-wild-challenge/
(b) S. R. Livingstone et al., "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," PLOS ONE, vol. 13, 2018

Motivation

Traditional Approaches

- Models **trained** on benchmark datasets enable **generalisation** across contexts and environments.
- **Yet,** generalisation comes **at the cost** of personalised learning.
- **Costly** to retrain and update models *on-the-fly.*

Personalisation towards Individual Expression

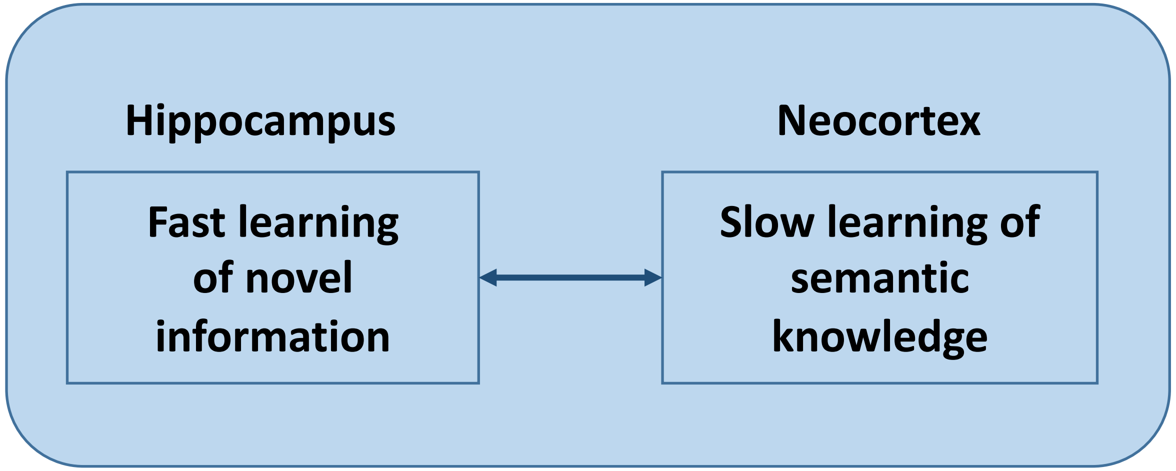
- Models that **continually learn** and **adapt** with each user.
- **Adaptation** with **new data** acquired during **real-world interactions** with users **without forgetting** the learnt knowledge.
- ***Continual Learning*** of Individual Facial Expressions to embed **personalisation** in models.

UNIVERSITY OF CAMBRIDGE

S. Li et al., "Deep facial expression recognition: A survey," CoRR, vol. abs/1804.08348, 2018.
https://marketingland.com/wp-content/ml-loads/2016/03/emotions_ss_1920.png

Complementary Learning Systems in the Human Brain

- Hippocampal and neocortical regions of the brain form a **complementary learning system**.
- Hippocampus forms an **episodic memory** for learning **novel information**.
- Neocortex forms a **semantic memory** by slowly **replaying** information from the hippocampus enabling **long-term retention**.



Growing Dual Memory (GDM)

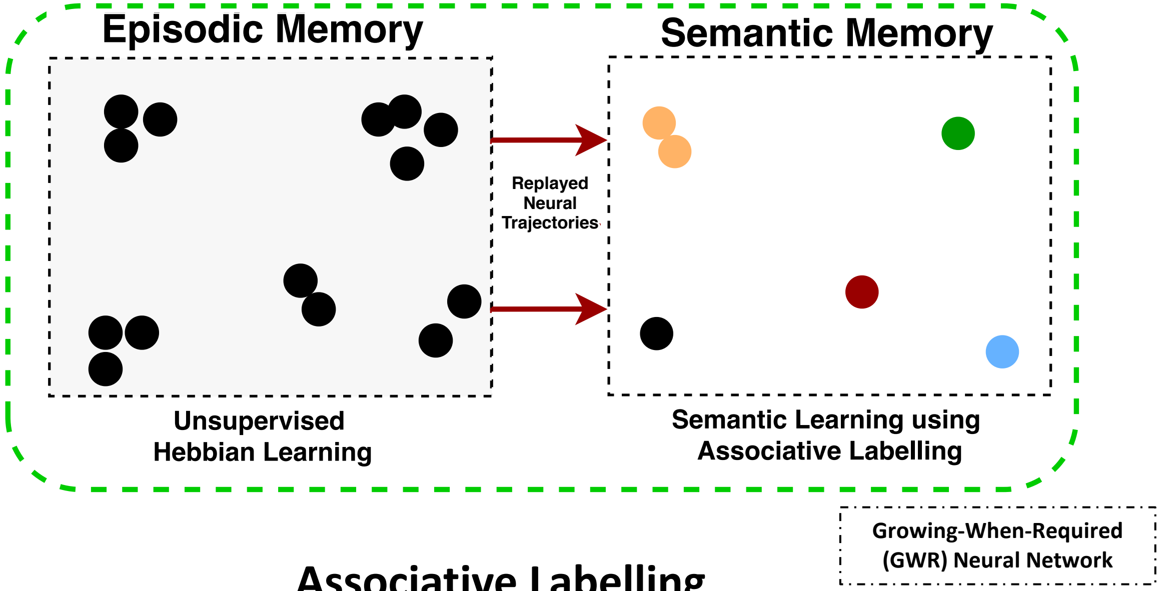
Episodic Memory learns ***non-overlapping*** representations of **novel** experiences.

- Distance-based similarity for **unsupervised Hebbian learning**.
- New neurons added **rapidly** to learn feature prototypes.

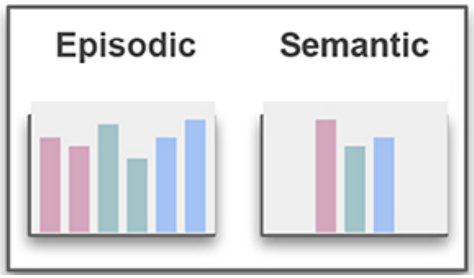
Semantic Memory learns compact ***overlapping*** representations that **generalise** across a particular class.

- Best Matching Neurons (BMUs) from episodic memory **replayed** to the semantic memory.
- ***Slow updation*** enables **overlapping** representations.
- Histogram frequency-based **associative labelling** for classification.

Pseudo-rehearsal to guard against forgetting by **replaying** trajectories of neural activations from the episodic memory.

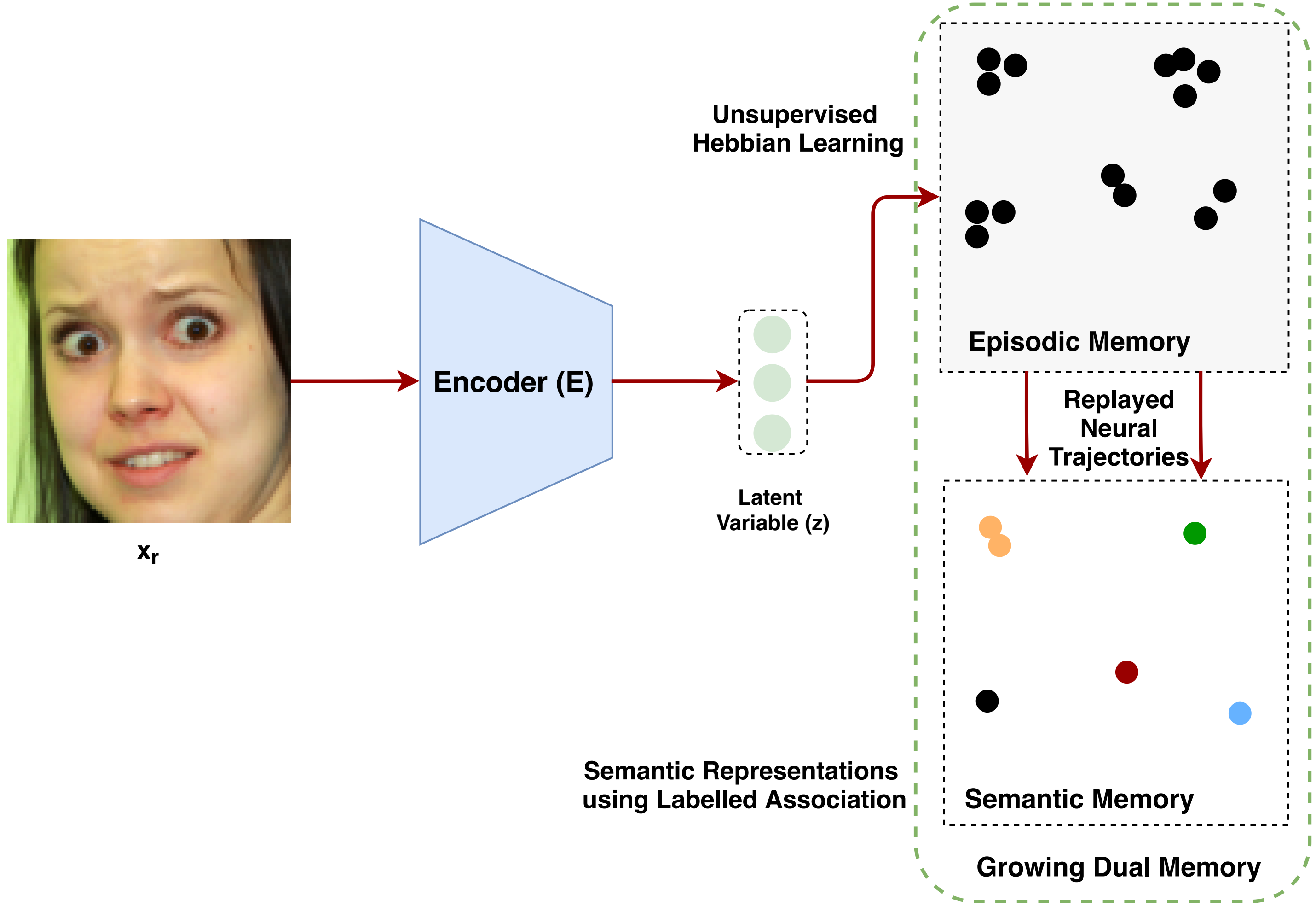


Associative Labelling



G.I. Parisi, et al. "Lifelong learning of spatiotemporal representations with dual-memory recurrent self-organization." Frontiers in neurorobotics 12 (2018): 78.

Continual Learning for Facial Expression Recognition

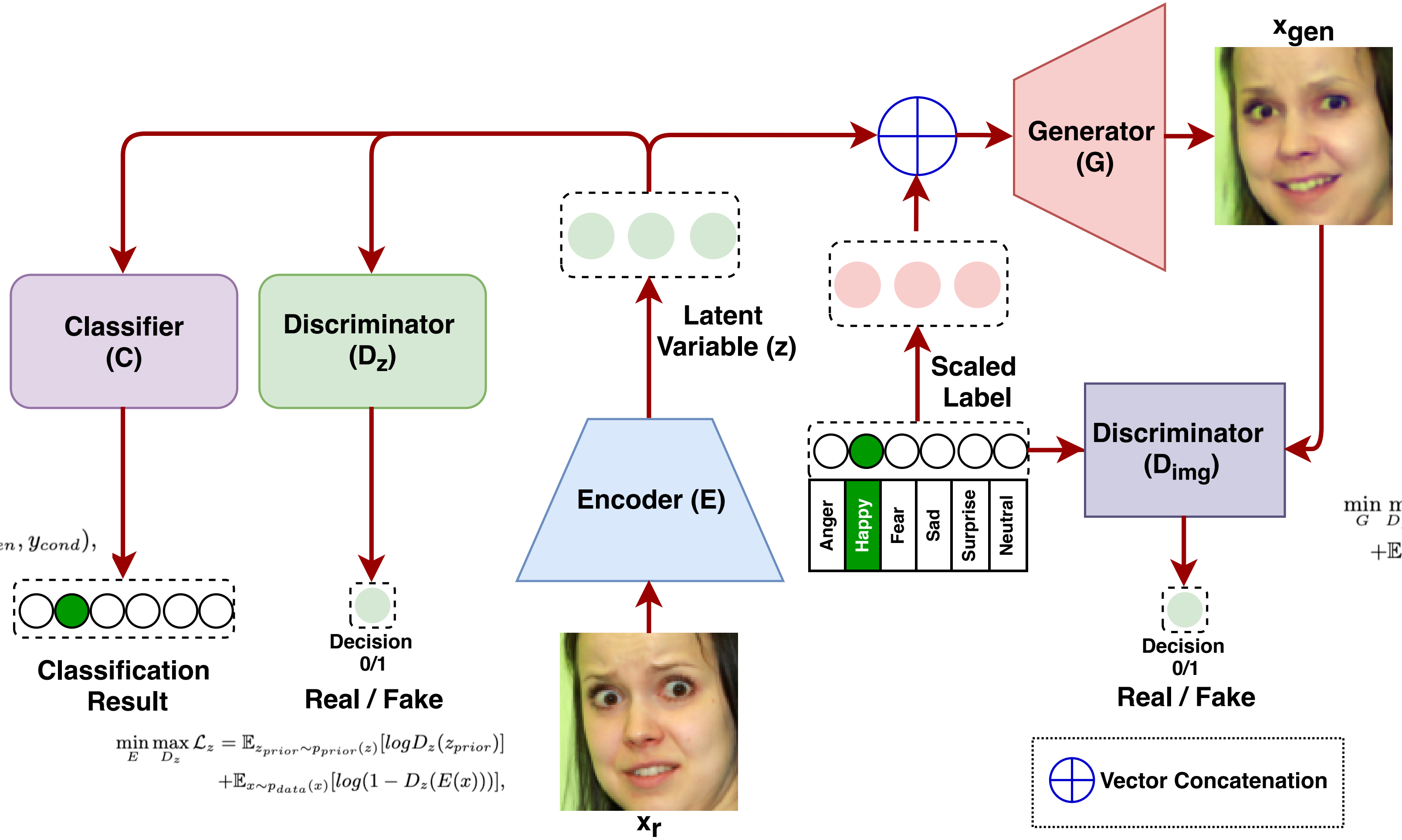


Challenges and Solutions

- Lack of **subject-specific data** acts as bottleneck for enabling **personalisation.**
- **Imagination** as a substitute to sensory input.
- **Simulated data** enables **individual and context-dependent adaptation**.
- *Adversarial Learning* to **imagine** additional data to augment learning.

Imagination Model

Conditional Adversarial Auto-Encoder (CAAE)



$$\min_{E,G} \mathcal{L}_{rec} = L_1(x_r, G(E(x_r), y)),$$

$$\min_{E,G} \mathcal{L}_{ID} = \sum_l L_1(\phi_l(G(E(x_r), y), \phi_l(x_r)),$$

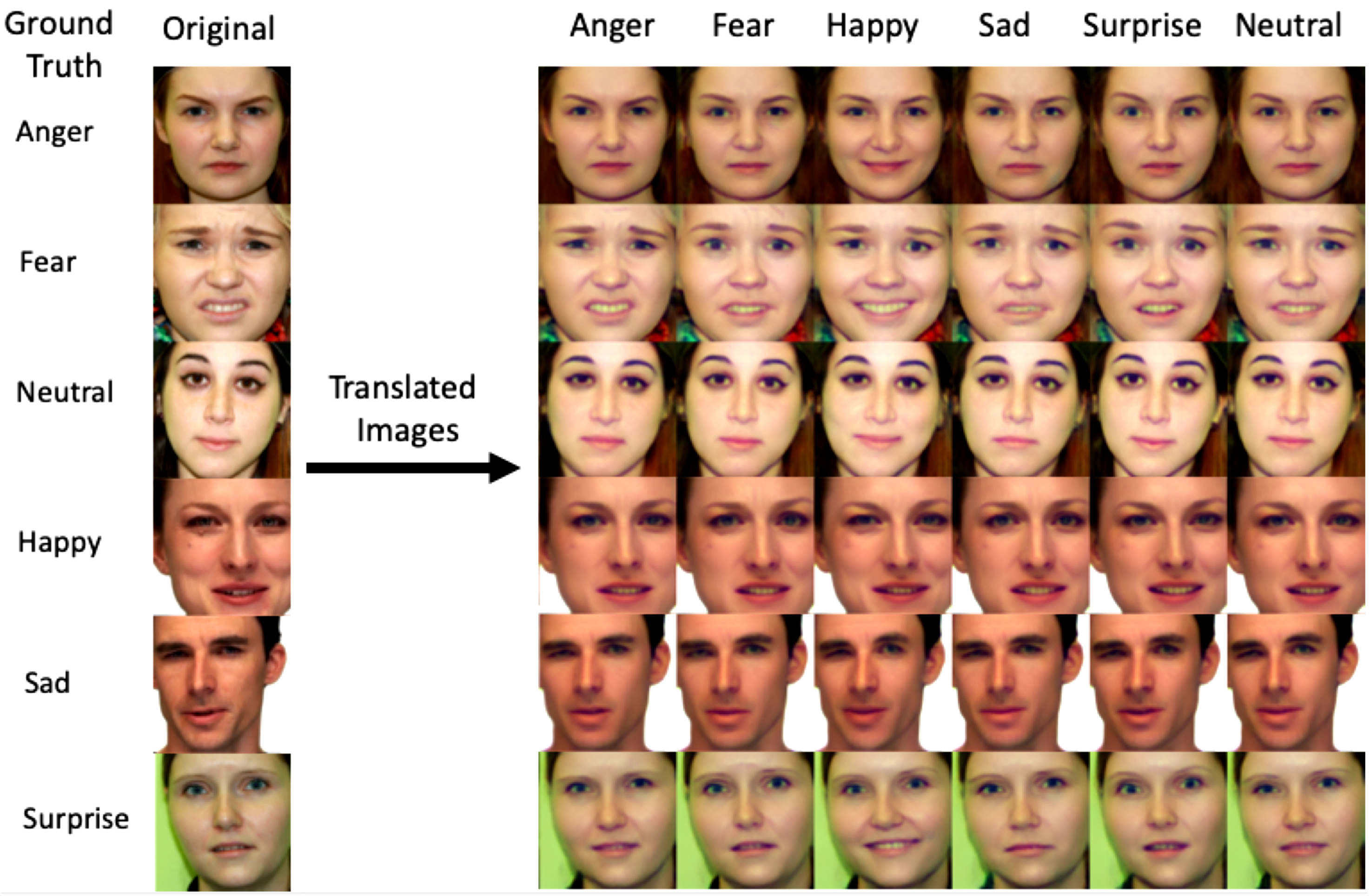
$$\min_{G} \max_{D_{Img}} \mathcal{L}_{img} = \mathbb{E}_{x,y\sim p_{data}(x,y)}[log(D_{Img}(x,y))] + \mathbb{E}_{x,y\sim p_{data}(x,y)}[log(1 - D_{Img}(G(E(x),y),y)],$$

$$\min_{\theta_C} \mathcal{L}_C = \phi(x_r, y_r) + \lambda_c \phi(x_{gen}, y_{cond}),$$

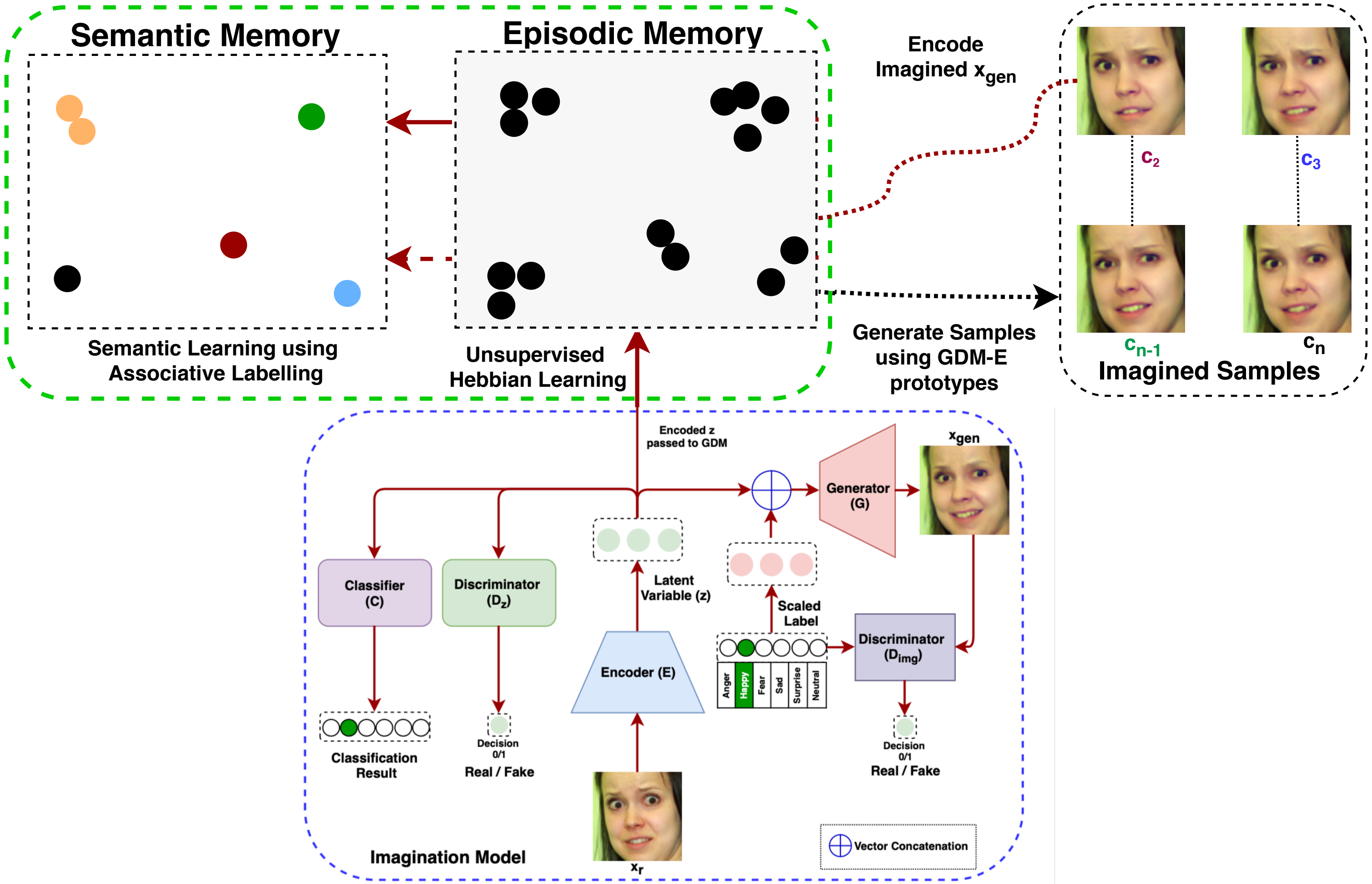
$$\min_{E} \max_{D_z} \mathcal{L}_z = \mathbb{E}_{z_{prior}\sim p_{prior}(z)}[log D_z(z_{prior})] + \mathbb{E}_{x\sim p_{data}(x)}[log(1 - D_z(E(x)))],$$

Imagination Model

Generated Images



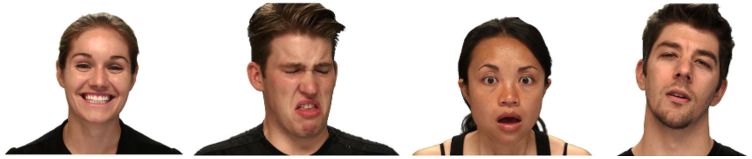
Continual Learning with Imagination for FER



Evaluation

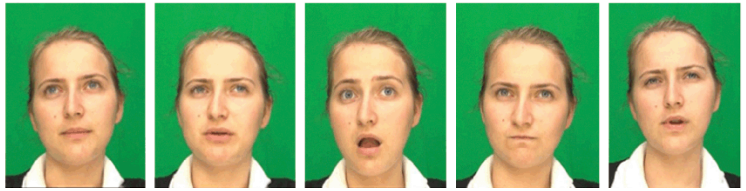
Datasets:

- Evaluating model performance on **three datasets** for *six classes*:
    - *Anger, Sadness, Happiness, Surprise* and *Fearful* and *Neutral.*
- **RAVDESS** (24 Subjects)



- **MMI** (10 Subjects)

- **BAUM-1 Spontaneous** (9 Subjects)



Comparisons:

- **Encoder with MLP-Baseline:**
    - Traditional Batch-Learning.
    - Model Retrained with each new class.
    - *Baseline* for traditional ML.
- **Encoder with GDM:**
    - Growing Dual Memory model **without** *pseudo-rehearsal*.
    - *Baseline* for CL.
- **Encoder with GDM + Replay:**
    - Growing Dual Memory model **with** *pseudo-rehearsal*.
    - *Explicit replay* of neural trajectories following seen classes.
- **Encoder with CLIFER (GDM + Imagination):**
    - Growing Dual Memory model **with** *imagination*.
    - *Implicit replay* of seen and unseen (imagined) classes.

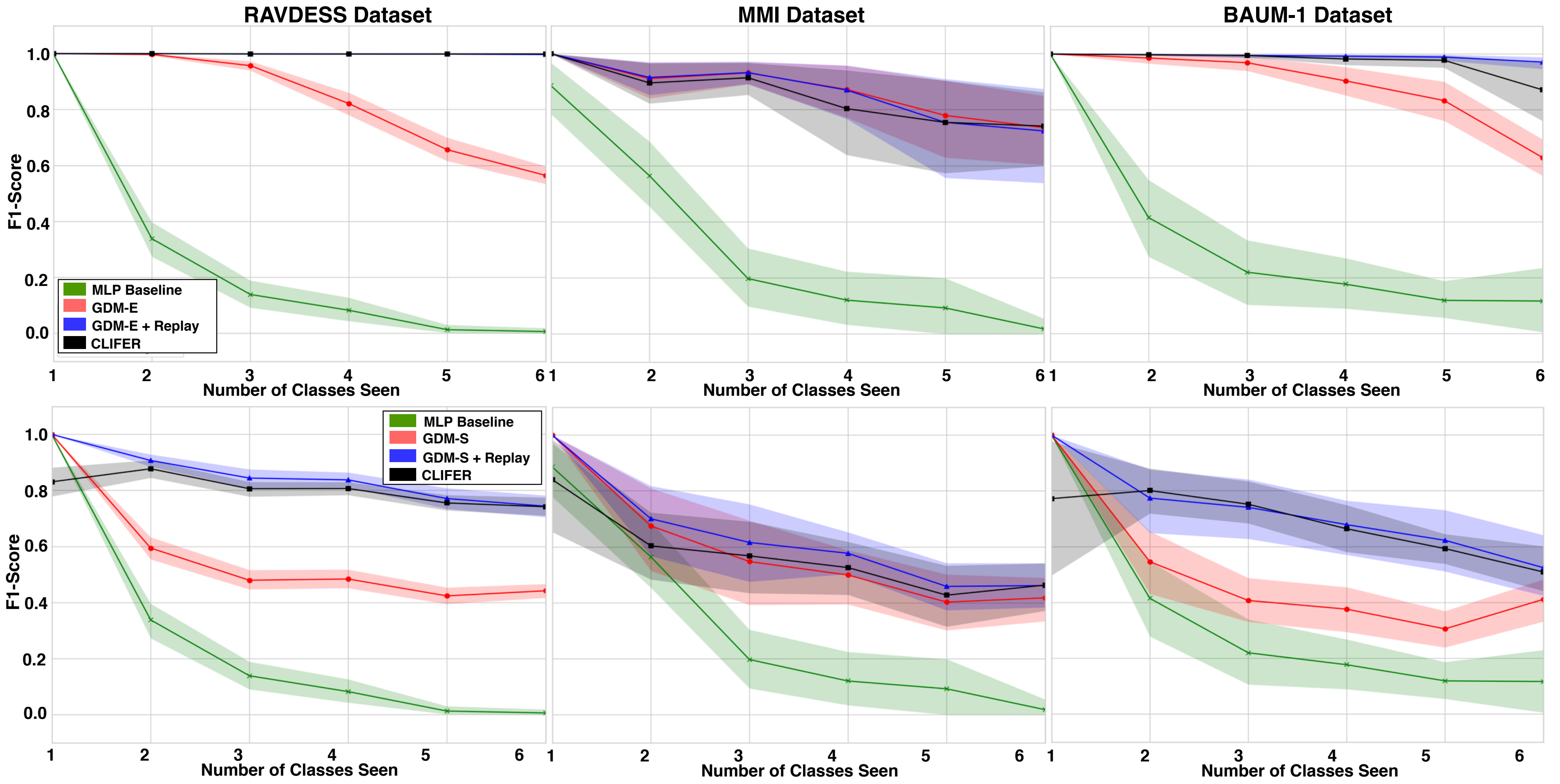
UNIVERSITY OF CAMBRIDGE

Experiment 1: Remembering Seen Expressions

Class Incremental learning training the models one class at a time for each subject.

After each new class, models evaluated on ***previously seen*** classes.

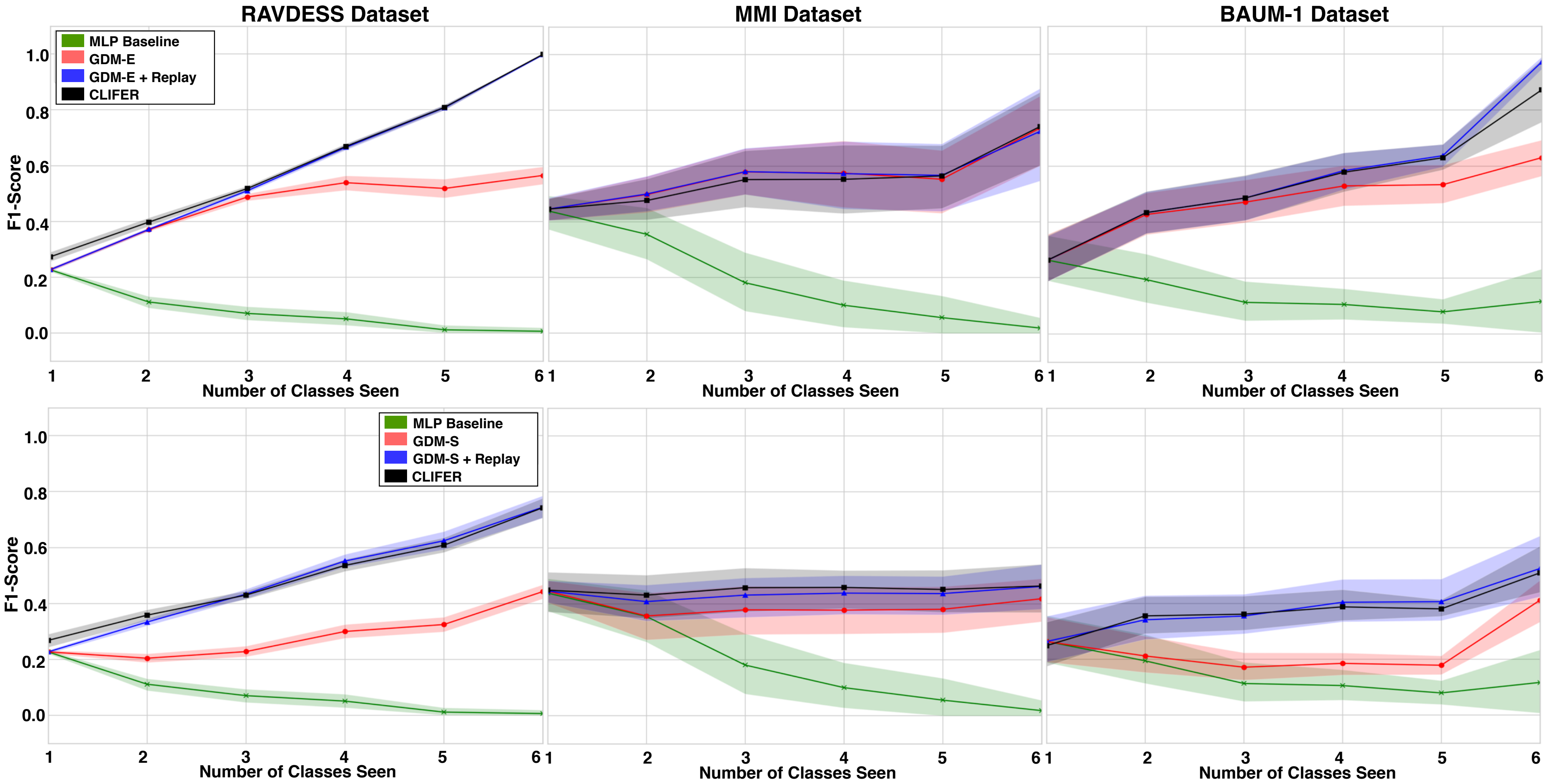
Mean F1-Score with **95% confidence** interval reported across all **subjects**.



Experiment 2: Adapting to New and Unseen Expressions

After each new class, models evaluated on *all* ***seen*** *and* ***unseen*** classes.

Mean F1-Score with **95% confidence** interval reported across all **subjects**.



Class Ordering Impacts Model Performance

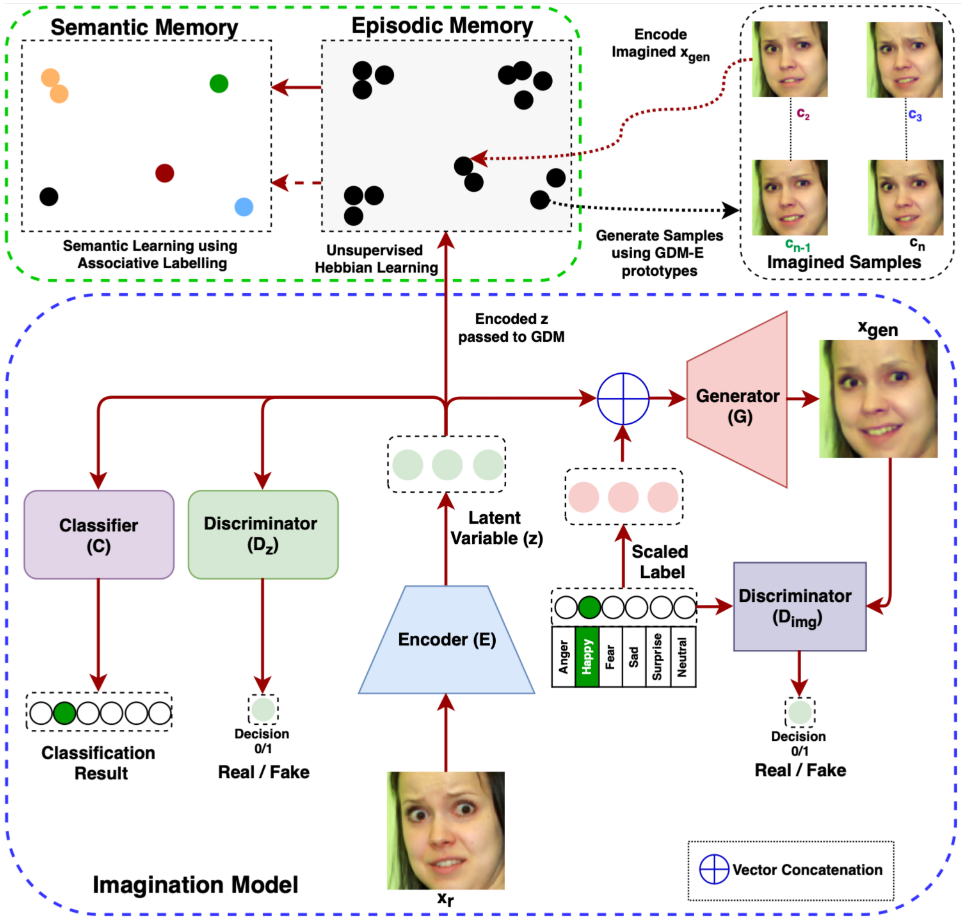
- Model performance **sensitive** to order of learning classes.
- CLIFER performance compared for **6 learning orders** starting with each class once and randomly selecting the others.
- Kruskal-Wallis H-tests report **significant *($p < 0.05$)*** difference in F1-scores at the beginning and end of learning for Experiment 2.
- Starting with ***neutral*** results in best model performance across all 3 datasets. **Order of learning** set to start with *Neutral*, followed by (randomly selected) *Happiness, Surprise, Anger, Fearful* and *Sadness*.
- ***Neutral*** seems to represent a **normative baseline** that allows **distinct feature prototypes** for subsequent classes.

F1-Score		
Dataset	**Episodic**	**Semantic**
RAVDESS	0.98 ± 0.01	0.75 ± 0.01
MMI	0.75 ± 0.07	0.46 ± 0.04
BAUM-1	0.87 ± 0.05	0.51 ± 0.04

CLIFER Performance

Conclusions

- We present a **novel framework** applying *Continual Learning with Imagination for Facial Expression Recognition*.
- Complementary Learning based solution **integrates new information** without interfering **with past knowledge**.
- **Imagination** as a critical tool for **simulating person-specific data** in real-world applications.
- **Order** of learning different expression classes **impacts** model performance where **starting neutral** improves future learning.



UNIVERSITY OF CAMBRIDGE
Department of Computer Science and Technology

15th IEEE International Conference on Automatic Face and Gesture Recognition

CLIFER: Continual Learning with Imagination for Facial Expression Recognition

Nikhil Churamani Hatice Gunes

Engineering and Physical Sciences Research Council