Energy-based Latent Aligner for Incremental Learning: Supplementary Material

K J Joseph[†‡] Salman Khan[‡⋆] Fahad Shahbaz Khan[‡⋄] Rao Muhammad Anwer[‡¶]
Vineeth N Balasubramanian[†]

[†]Indian Institute of Technology Hyderabad, India [‡]Mohamed bin Zayed University of AI, UAE
[⋆]Australian National University, Australia [⋄]Linköping University, Sweden [¶]Aalto University, Finland

In this supplementary material, we provide additional details and experimental analysis regarding the behaviour of proposed latent alignment approach (ELI). They are:

- An illustration for the adaptation process of latent representation with ELI. (Sec. S1)
- Effect of using *mixup* for data augmentation. (Sec. S2)
- Details regarding the submitted codebase. (Sec. S3)
- Comments on the broader societal impacts. (Sec. S4)
- Qualitative results on incremental detection. (Sec. S5)
- A summary of notations used in the paper. (Sec. S6)

S1. Recognizing Important Latents Implicitly

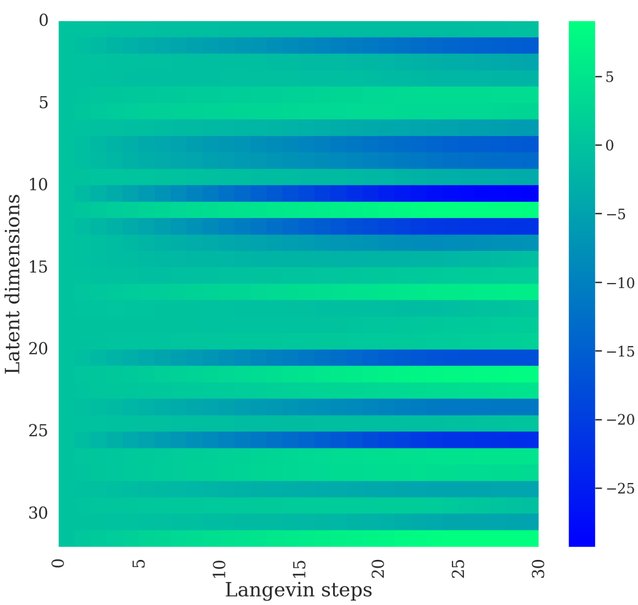

Figure S1. Each row i shows how i^{th} latent dimension is updated by ELI. We see that different dimensions have different degrees of change, which is implicitly decided by our energy-based model.

Fig. S1 shows how each latent dimension of a 32 dimensional latent vector (y-axis) gets adapted in each Langevin iteration (x-axis). For an initial latent representation $\mathbf{z}_0$, each column shows the difference from its aligned version from the i^{th} Langevin step: $\mathbf{z}_i - \mathbf{z}_0$. We consider MNIST experiment (Sec. 3.3) for this illustration. Our proposed latent aligner is able to implicitly identify which latent dimension is important to be preserved or modified. This characteristic is difficult to achieve in alternate regularization methods like distillation, which gives equal weightage to each dimension. We can see that the specialization happens within a few number of iterations, similar to the results in Tab. 3. Visualization in Fig. S1 is using a trained EBM. `latents.mp4` file (attached in supplementary) shows how the latent representations change as the EBM is learned.

S2. Augmenting Data with *mixup*

As detailed in Sec. 3.2, we use datapoints sampled form the current task distribution to learn the energy-based model $\mathbf{x}_i \sim p_{data}^{\mathcal{T}_t}$. Here we use *mixup*, an augmentation technique introduced by Zhang *et al.* [3], where each datapoint is modified as $\hat{\mathbf{x}} = \lambda \mathbf{x}_i + (1-\lambda)\mathbf{x}_j$, s.t. $\lambda \sim Beta(\alpha, \alpha)$, and report the results in Tab. S1. In these experiments with incremental CIFAR-100, we see that using *mixup* does not enhance performance, even with different values of α. This is because the EBM is a small two layer network which is not prone to overfitting, and can perform well even without this extra augmentation.

Table S1. The performance of EBM is comparable with and without using *mixup* augmentation as the EBM network is small.

α	5 Tasks	10 Tasks	25 Tasks
Without *mixup* [3]	63.68	58.92	54.00
0.1	63.67	58.85	54.01
0.3	63.53	58.81	53.85
0.5	63.54	58.79	53.88
1.0	63.44	58.53	53.83

S3. Code

We enclose two Jupyter notebooks which contain code for the 32 dimensional and 512 dimensional MNIST experiments explained in Sec. 3.3. They are titled `ELI.ipynb` and `ELI_512.ipynb` respectively. Importantly, the Python

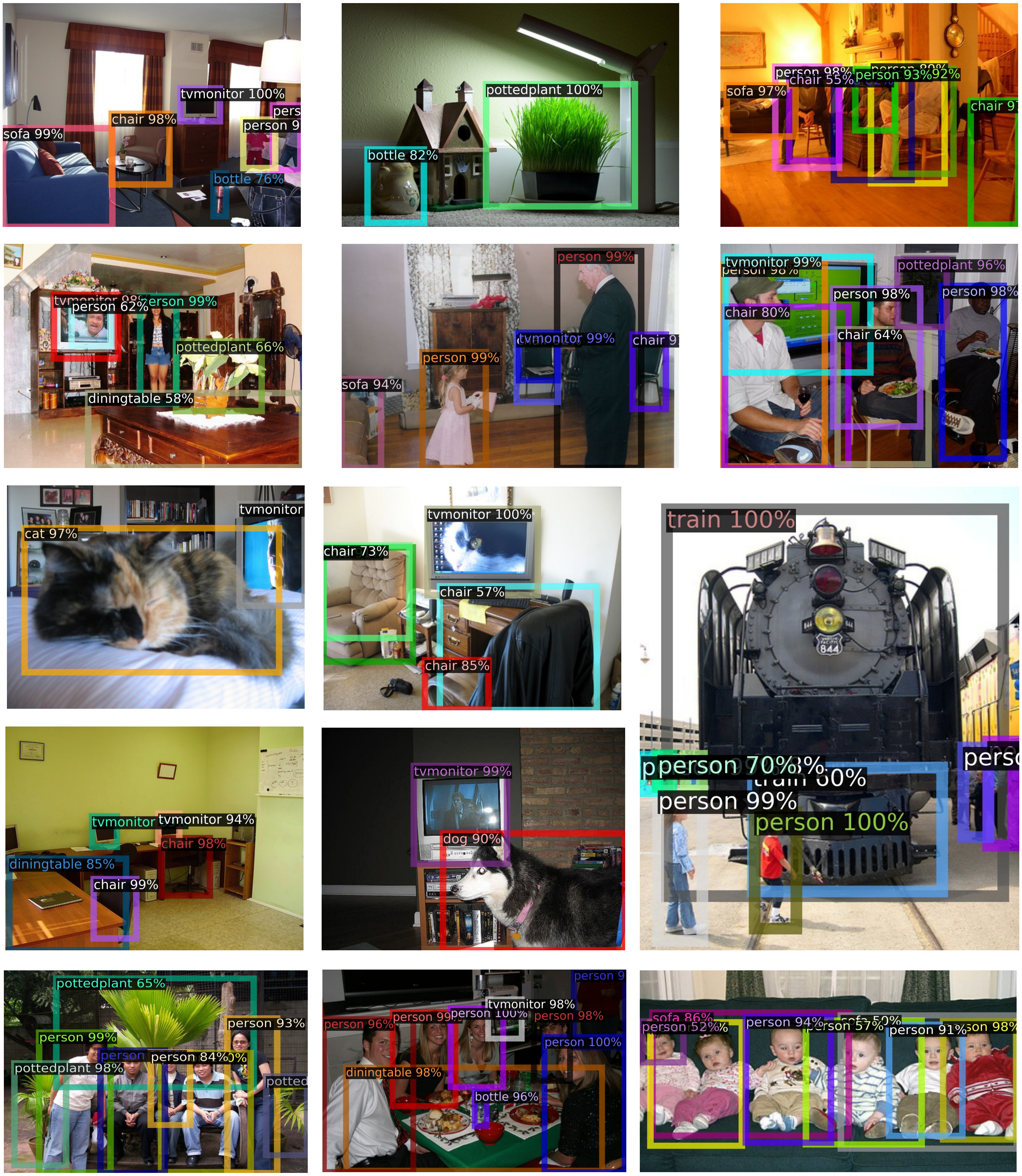

Figure S2. Qualitative results of incremental Object Detection. We consider the $10+5$ setting on Pascal VOC, where instances of `plant`, `sheep`, `sofa`, `train` and `tvmonitor` are added to a detector trained on the rest of the classes.

class `EBMAligner` is used with minimal modification for all other large scale incremental classification and object detection experiments in the paper. Our codes and models are available at https://github.com/JosephKJ/ELI.

S4. Broader Impact

When a model incrementally learns without forgetting, an equivalently important desiderata would be to selectively forget, in adherence to any privacy or legislative reasons. Such an unlearning can be possible by treating such instances as out-of-distribution samples, however, a dedicated treatment of the same is beyond the current scope of our work. Our current work aims to reduce the catastrophic forgetting and interference while learning continually, and to the best of our knowledge, our methodology does not have any detrimental social impacts that make us different from other research efforts geared in this direction.

S5. Qualitative Results

In Figure S2, we show more qualitative results for incremental Object Detection in the $15 + 5$ setting with Pascal VOC dataset [1]. Instances of `plant`, `sheep`, `sofa`, `train` and `tvmonitor` are added to a detector trained on the rest. The considerable improvement of ELI over the state-of-the-art-method [2] as shown in Tab. 2, is due to the implicit latent space regularization that ELI offers. To the best of our knowledge, ELI is the first method that adds latent space regularization to large scale incremental object detection models.

S6. Summary of Notations

For clarity, Tab. S2 summarizes the main notations used in our paper along with their concise description.

Table S2. To enhance readability, this table summarises the notations used in the manuscript, along with their meaning.

Notation	Stands for
τ_i	i^{th} task
$\mathbf{x} \in \tau_i$	Image from the i^{th} task
$\mathcal{T}_t = \{\tau_1, \tau_2, \cdots, \tau_t\}$	Continuum or set of tasks seen until time t
$\mathbf{x} \in \mathcal{T}_t$	Image from any of the task in $\mathcal{T}_t$
$\mathcal{M}^{\mathcal{T}_t}$	Model trained until time t
$\mathcal{F}_{\boldsymbol{\theta}}^{\mathcal{T}_t}$	Feature extractor of $\mathcal{M}^{\mathcal{T}_t}$
$\mathcal{F}_{\boldsymbol{\phi}}^{\mathcal{T}_t}$	Task specific part of $\mathcal{M}^{\mathcal{T}_t}$
$\mathbf{z}^{\mathcal{T}_t}$	Latent representation from $\mathcal{F}_{\boldsymbol{\theta}}^{\mathcal{T}_t}$
$p_{data}^{\tau_t}$	Data distribution of task τ_t
$(\mathbf{x}_i^{\tau_t}, y_i^{\tau_t}) \sim p_{data}^{\tau_t}$	Samples from $p_{data}^{\tau_t}$

References

[1] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. S3

[2] KJ Joseph, Jathushan Rajasegaran, Salman Khan, Fahad Khan, and Vineeth N Balasubramanian. Incremental object detection via meta-learning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Nov 2021. S3

[3] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *ICLR*, 2018. S1