Expert Gate: Lifelong Learning with a Network of Experts (Supplementary Materials)

1 Comparison of different autoencoder design choices

We show the effect of using different design choices for our autoencoder gate. As a test case, we use the 3 sequential learning tasks. Namely, Scenes, Birds and Flowers. We consider the following alternatives:

- *Linear autoencoder:* no use of nonlinear activation functions. The loss function is the euclidean distance. This setup learns the same subspace as PCA.
- *No standardization:* the same structure of our autoencoder gate but without the standardization of the input. Also, no fine-tuning is performed here.
- *Sigmoid activation:* in our design choice, we use a ReLU activation function for the encoding layer. In this baseline, we use the sigmoid activation function for the encoding as well as decoding layer.
- *No ImageNet fine-tuning:* we show the performance of our autoencoder gate without the initialization of the ImageNet pretrained autoencoder.
- *Expert Gate:* our autoencoder gate with the full design: standardization step, ReLU hidden activation and ImageNet fine-tuning.

For all the different alternatives, we use the adaptive gradient in the training of the autoencoder. Table 1 shows the classification accuracy for the task labeling problem achieved by each of the different choices and our Expert Gate autoencoder. It can be noticed that the linear gate (*Linear Autoencoder*) fails to recognize the examples from the Flowers dataset – the linearly learned subspace might not be different from the subspace learned for Birds (due to the visual similarity). There is a slight relative improvement between the gate with the standardization (*ReLU activation function* and *Sigmoid activation function*) and without (*No standardization*). A small improvement can also be seen when using ReLU over the sigmoid activation function. Lastly, our design choice along with finetuning after the autoencoder learned on ImageNet achieves the highest accuracy in recognizing the three different tasks.

Table 1: Comparison of different autoencoder designs: classification accuracy of the autoencoders for the sequential learning of 3 image classification tasks.

Method	Scenes	Birds	Flowers	avg
Linear Autoencoder	93.6	98.3	35.4	75.8
No standardization	97.4	97.7	97.4	97.5
Sigmoid activation function	97.3	98.4	97.4	97.7
ReLU activation function	97.6	98.4	97.4	97.8
Expert Gate	99.4	99.2	99.2	99.3

2 Confusion Examples

Here we consider the six image classification tasks, i.e. Scenes, Birds, Flowers, Cars, Aircrafts and Actions. As a continuation of our gate analysis, we show here more qualitative examples. In Figure 1, you can find for each task: an example that has been assigned mistakenly to one of the other tasks. Note that for some cases, the examples shown are the only mistakes made by our Expert Gate.

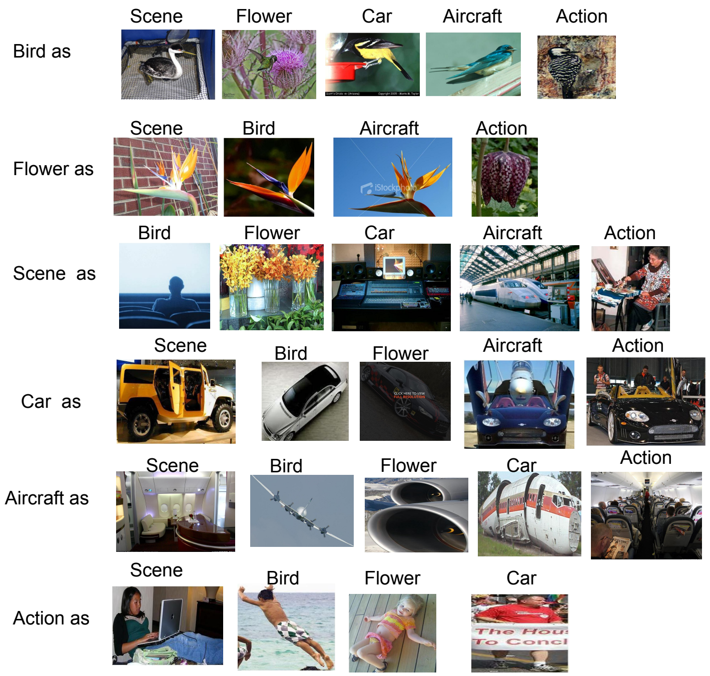

Figure 1: Confusion Cases.

Most of these mistakes are explainable and due to one of the following reasons:

- the image contains objects from two different tasks and our Expert Gate has to choose one of them. This can be handled by allowing more than

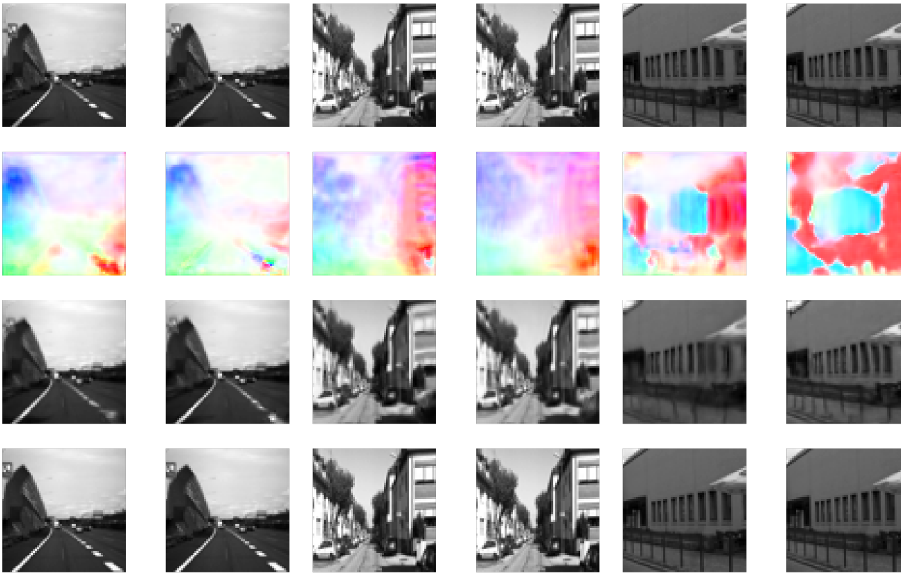

Figure 2: Video prediction on the Highway, Residential and City datasets from left to right (two columns for each dataset) using sequential fine-tuning (first column) and our Expert Gate (second column).

one expert to be activated as we have shown in the paper.

- the image is an outlier w.r.t. its own task dataset. For example, only a small part of the object appears which means it is not a useful example or it could even harm the classifier if used in the training phase. This sheds light on another potential use of our gate, i.e. to detect outliers. In fact, the autoencoder represents the distribution of each task data. Thus an outlier that is in the long tail of the task distribution will have a higher reconstruction error. This has the potential of being used in cleaning the annotations of each new task data.

- objects from one task that look similar to the images of another task.

3 Video Prediction: Qualitative Results

Figure 2 shows qualitative results from the video prediction experiments using sequential fine-tuning and Expert Gate. The first row shows the first in a sequence of 3 images input from the highway, residential and city datasets, to sequential fine-tuning and Expert Gate respectively. The second row has a visualization of the video prediction filters for the two methods in the 3 datasets. The visualization is done in the same way as in [1], and shows the output of the filters as a flow field indicating the predicted motion in the images. The third row contains the last of the three images predicted by the 2 systems, again for the 3 datasets. Ground truth predictions are shown on the last row. It can be seen that Expert Gate gives consistently superior qualitative results in the 3 datasets.

References

[1] De Brabandere, B., Jia, X., Tuytelaars, T., Van Gool, L.: Dynamic filter networks. arXiv preprint arXiv:1605.09673 (2016)