Summary

Bardia Mojra
September 22, 2021
Seminar on Continual Learning
Robotic Vision Lab

Unified Probabilistic Deep Continual Learning through Generative Replay and Open Set Recognition

In this paper, the authors introduce a unified probabilistic model to tackle the open set recognition problem while mitigating catastrophic forgetting in deep continual learning. They deploy a Bayesian variational auto-encoder, [1], to model joint probability distributions posterior for the auto-decoder and linear classifier. Their approach combines a joint-probability encoder with a generative replay model and a linear classifier for regularization and final classification, respectively. ELBO and EVT, [2] [3], are cleverly deployed to put a tight bound on learned parameter latent variables in high density regions. Inspired by [4], they deploy the following loss function,

$$\mathcal{L}\left(x^{(n)}, y^{(n)}; \theta\phi, \xi\right) = -\beta KL\left(q_\theta(z|x^{(n)})\|p(z)\right) + \mathbb{E}_{q_\theta(z|x^{(n)})}\left[\log p_\phi(x^{(n)}|z) + \log p_\xi(y^{(n)}|z)\right] \tag{1}$$

where θ represents the shared encoder parameters. ϕ and ξ represent parameters for the decoder and the linear classifier. The joint probabilistic encoder learns to encode latent variable vector z with an assumed unit Gaussian distribution. Per mentioned variational inference methods, the auto-encoder, θ, learns joint probability distribution for latent variable z with both input and output, represented by $p_\phi(x, z)$ and $p_\xi(y, z)$ respectively. Figure 1 represents the unified probabilistic model presented in this paper,

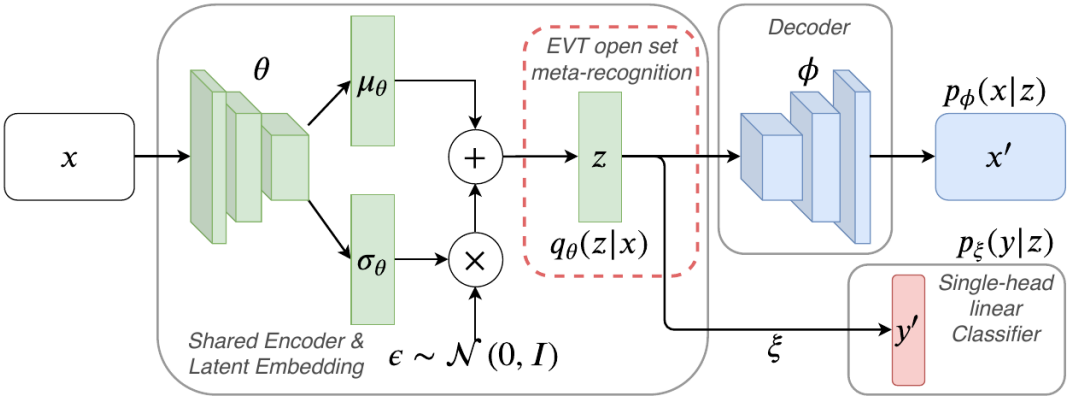
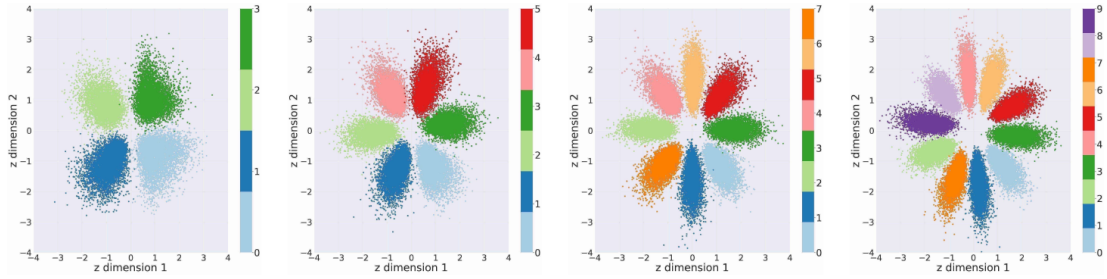


Figure 1: Joint continual learning model consisting of a shared probabilistic encoder $q_\theta(z|x)$, probabilistic decoder $p_\phi(x, z)$ and probabilistic classifier $p_{xi}(y, z)$. For open set recognition and generative replay with outlier rejection, EVT based bounds on the basis of the approximate posterior are established.

This is particularly important to understand. Now, we have two joint probabilistic distributions consisting of three separate distributions, x, y, and z. Out of the three separate distributions, two are known, x and y which, represent the training set input and ground truth labels. Variational inference methods takes advantage such problem setting by assuming that a joint probability distribution of a large-enough sample set, 1) follow the same or very similar distribution of the population. 2) The non-normalized sample distribution is always smaller than the population distribution. 3) And the sample distribution is directly proportional and representative of the population distribution. On a side note, estimating the population distribution from a sample distribution through iterative means is referred to as KL-divergence and is used when deploying variational inference methods. Moreover, variational inference decomposes complex distributions into individually recognizable distributions, z latent space variable vector. Thus, we can use a variational inference based auto-encoders to approximate the aggregate posterior as the following,

$$q_{\theta,t}(z) = \mathbb{E}_{p_{\tilde{D}_t}(\tilde{x})}[q_{\theta,t}(z|\tilde{x})] \approx \frac{1}{\tilde{N}_t} \sum_{n=1}^{\tilde{N}_t} q_{\theta,t}(z|\tilde{x}^{(n)}) \tag{2}$$

Figure 2 depicts latent space variable projections onto $z_1 \times z_2$ phase plane.



(a) 4-classes (b) 6-classes (c) 8-classes (d) 10-classes

Figure 2: 2-D latent space visualization for continually learned MNIST.

The following figures depict increased performance after rejecting outliers.

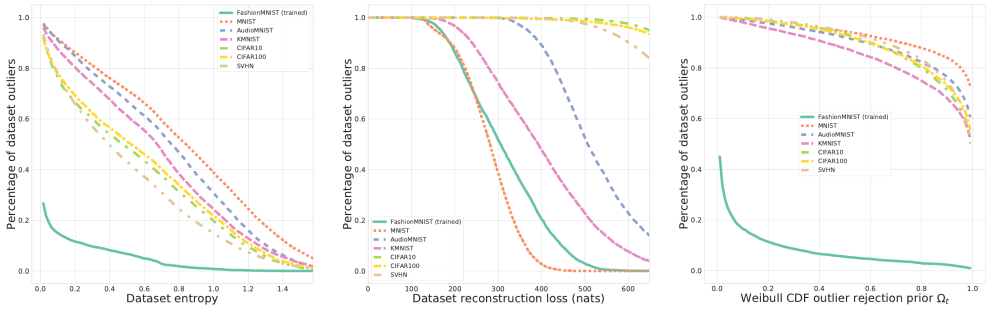


Figure 3: Trained Fashion MNIST OCDVAE evaluated on unknown datasets. All metrics are averaged over 100 approximate posterior samples per data point. (Left) Classifier entropy values are insufficient to separate most of unknown from the known task's test data. (Center) Reconstruction loss allows for a partial distinction. (Right) Our posterior based open set recognition considers the large majority of unknown data as statistical outliers across a wide range of rejection priors Ω_t .

References

[1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[2] M. D. Hoffman and M. J. Johnson, "Elbo surgery: yet another way to carve up the variational evidence lower bound," in *Workshop in Advances in Approximate Bayesian Inference, NIPS*, vol. 1, 2016.

[3] A. Bendale and T. Boult, "Reliable posterior probability estimation for streaming face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 56–63, 2014.

[4] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," 2016.