NestedVAE: Supplementary Material

Matthew J. Vowels
m.j.vowels@surrey.ac.uk

Necati Cihan Camgoz
n.camgoz@surrey.ac.uk

Richard Bowden
r.bowden@surrey.ac.uk

Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, UK

1. Overview of Supplementary Material

This material supplements the paper 'NestedVAE: Isolating common factors via weak supervision.' Firstly, we present the algorithm for training NestedVAE in Algorithm 1. Secondly, we present additional results for biological sex prediction on the UTKFace dataset. Finally, we present architectures used for the MNIST and UTKFace experiments in Figure 3.

Algorithm 1 NESTEDVAE training procedure.

Input: Image pairs $\mathbf{x}_i \sim X_1$ and $\mathbf{x}_j \sim X_2$ $i, j \in [1, N]$, learning rate α, loss weights γ, λ,
Output: NestedVAE $\{\theta_1, \phi_1, \theta_2, \phi_2\}$
Initialisation :
Random init. [1] $\longrightarrow \{\theta_1, \phi_1, \theta_2, \phi_2\}$
Training
for $s = 0 : N/$batch size **do**
  Sample batch $\mathbf{x}_{i,j}$
  Outer VAE computations :
  Compute batch $\{\boldsymbol{\mu}_{i,j}, \boldsymbol{\sigma}_{i,j}\} = \mathrm{Enc}(\mathbf{x}_{i,j})$
  Sample batch $\mathbf{z}_{i,j} \sim N(\boldsymbol{\mu}_{i,j}, \boldsymbol{\sigma}_{i,j})$
  Compute batch $\hat{\mathbf{x}}_{i,j} = \mathrm{Dec}(\mathbf{z}_{i,j})$
  Compute batch loss $\mathrm{MSE}(\hat{\mathbf{x}}_{i,j}, \mathbf{x}_{i,j})$
  Compute batch loss β KL$\big[(\boldsymbol{\mu}_{i,j}, \boldsymbol{\sigma}_{i,j}), N(0, I)\big]$
  Nested VAE computations :
  Compute batch $\{\boldsymbol{\mu}_{Nest-i}, \boldsymbol{\sigma}_{Nest-i}\} = \mathrm{Enc}(\boldsymbol{\mu}_i)$
  Sample batch $\mathbf{z}_s \sim N(\boldsymbol{\mu}_{Nest-i}, \boldsymbol{\sigma}_{Nest-i})$
  Compute batch $\hat{\boldsymbol{\mu}}_i = \mathrm{Dec}(\mathbf{z}_s)$
  Compute batch loss $\mathrm{MSE}(\hat{\boldsymbol{\mu}}_i, \boldsymbol{\mu}_j)$
  Compute batch loss $\beta_{Nest}\mathrm{KL}\big[(\hat{\boldsymbol{\mu}}_{Nest-i}, \boldsymbol{\sigma}_{Nest-i}), N(0, I)\big]$
  Combine :
  Weight VAE and NestedVAE losses by γ, λ respectively and sum.
  Backpropagate gradients and update w/ Adam [2] and learning rate $= \alpha$
end for

2. UTKFace

The UTKFace dataset [3] evaluation procedure is described in the main paper. The results for F1-score for males across race and females across race are shown in Figures 1 and 2, respectively. The complete F1-scores are shown in Table 1. It can be seen that NestedVAE significantly outperforms alternatives in all cases, and extracts the common factors for sex.

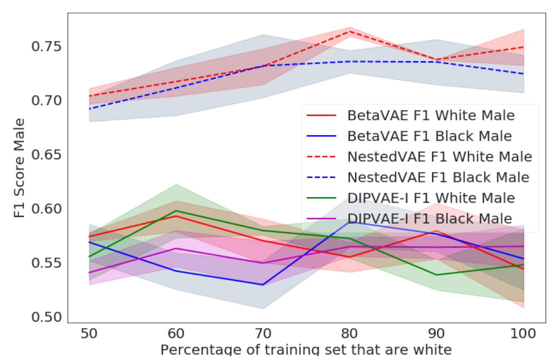

Figure 1. F1 scores for prediction of male sex using embeddings from models trained on datasets with varying proportions of white and black individuals. Best viewed in color.

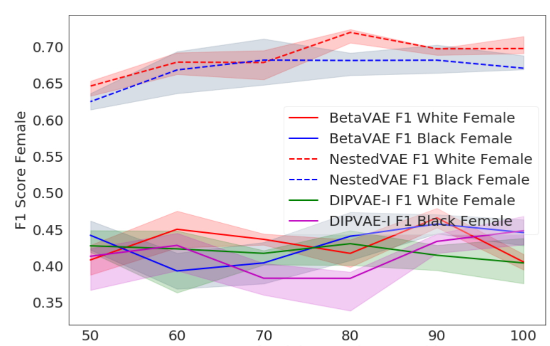

Figure 2. F1 scores for prediction of female sex using embeddings from models trained on datasets with varying proportions of white and black individuals.

	Percent White	Black Female	Black Male	White Female	White Male
β-**VAE**	100	0.446±0.018	0.553±0.029	0.405±0.011	0.544±0.036
	90	0.457±0.013	0.576±0.016	0.466±0.013	0.579±0.026
	80	0.441±0.033	0.587±0.023	0.417±0.019	0.555±0.014
	70	0.404±0.028	0.529±0.022	0.436±0.008	0.570±0.020
	60	0.393±0.025	0.542±0.017	0.450±0.025	0.593±0.014
	50	0.442±0.020	0.568±0.017	0.408±0.020	0.574±0.004
DIPVAE-I	100	0.448±0.019	0.565±0.020	0.404±0.034	0.548±0.020
	90	0.434±0.016	0.564±0.010	0.415±0.014	0.538±0.010
	80	0.383±0.044	0.564±0.009	0.430±0.018	0.572±0.009
	70	0.383±0.023	0.549±0.021	0.417±0.004	0.579±0.021
	60	0.428±0.034	0.564±0.017	0.423±0.025	0.597±0.017
	50	0.413±0.046	0.540±0.011	0.428±0.021	0.555±0.011
NestedVAE (ours)	100	**0.671±0.002**	**0.724±0.017**	**0.698±0.007**	**0.749±0.017**
	90	**0.682±0.018**	**0.735±0.021**	**0.698±0.008**	**0.737±0.001**
	80	**0.682±0.020**	**0.735±0.010**	**0.720±0.014**	**0.763±0.004**
	70	**0.682±0.034**	**0.731±0.029**	**0.679±0.023**	**0.731±0.017**
	60	**0.669±0.032**	**0.711±0.025**	**0.679±0.016**	**0.717±0.013**
	50	**0.625±0.011**	**0.692±0.012**	**0.646±0.013**	**0.703±0.007**

Table 1. F1 score results for classifier performance on black females, black males, white females, and white males using embeddings from models trained on data varying in the proportion of white individuals. Results demonstrate superior classification performance with NestedVAE. Best results are shown in **bold**.

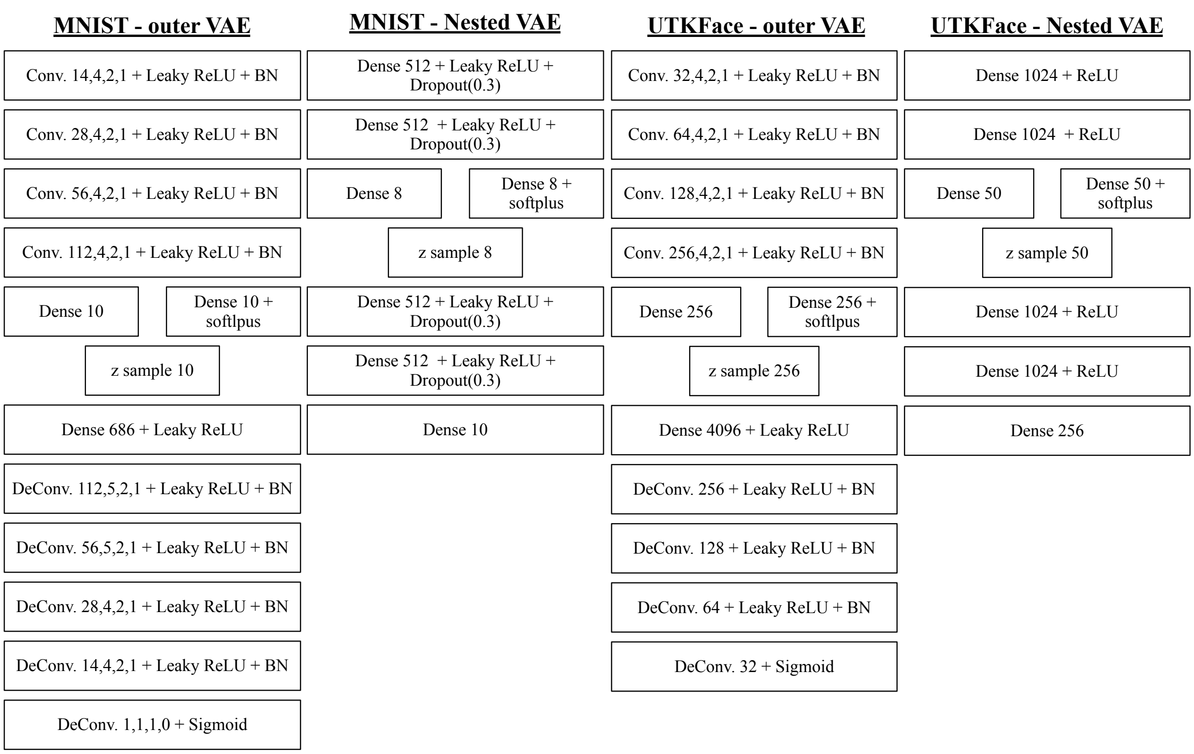

Figure 3. Network architectures used for the reported experiments on rotated MNIST and UTKFace datasets.

References

[1] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 2010.

[2] D. P. Kingma and J. L. Ba. Adam: a method for stochastic optimization. *arXiv:1412.6980v9*, 2017.

[3] Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. *arXiv:1702.08423*, 2017.