In simple terms, the sole purpose of validation set is to ensure that your model is learning as it is supposed to. So you must have training and validation from the same set.

Whereas test set is used to check how well your model generalize and how well it works on real-life data OR the data which it hasn’t seen.

So you can have test data from the same dataset through which your model learned OR get data from some new dataset OR combine and shuffle both. I’d prefer the last one.

they fall short dueto: a) reliance on fully-labelled data, and b) the inability to generate novel datafrom the abstractions.

The former is a fairly onerous requirement, particularly when dealing withreal-world visual data, as it requires many hours of human-annotator time andeffort to collect.

Such generative modelling, in contrast to discriminative modelling, en-ables ananalysis-by-synthesisapproach to human-body analysis, where one cangenerate images of humans in combinations of poses and clothing unseen duringtraining.

However, DGMs introduce a new problem – the learnt abstractions, or la-tent variables, are nothuman-interpretable. This lack of interpretability is a by-product of the unsupervised learning of representations from data. The learntlatent variables, typically represented as some smooth high-dimensional mani-fold, do not have consistent semantic meaning – different sub-spaces in thismanifold can encode arbitrary variations in the data. This is particularly un-suitable for our purposes as we would like to view and manipulate the latentvariables,e.g.the body pose.

n order to ameliorate the aforementioned issue, while still eschewing relianceon fully-labelled data, we rely on thestructured semi-supervisedvariational au-

toencoder (VAE) framework [17, 32]. Here, the model structure is assumed to bepartially specified, with consistent semantics imposed on some interpretable sub-set of the latent variables (e.g.pose), and the rest is left to be non-interpretable,although referred by us here asappearance. Weak (semi) supervision acts as ameans to constrain the pose latent variables to actually encode the pose.

hisgives us the full complement of desirable features, allowing a) semi-supervisedlearning, relaxing the need for labelled data, b) generative modelling throughstochastic computation graphs [28], and c) interpretable subset of latent vari-ables defined through model structure.

n summary, our main contributions are: i) a real-world application of struc-tured semi-supervised deep generative model of natural images, separating posefrom appearance in the analysis of the human body; ii) a quantitative and qual-itative evaluation of the generative capabilities of such model; and iii) a demon-stration of its utility in performing semi-supervised pose estimation and indirectpose-transfer.

The commonly used3D datasets [12, 24] were captured by mocap systems incontrolled lab environments. Deep neural networks [13, 33]trained on these datasets do not generalize well to other en-vironments, such as in the wild.