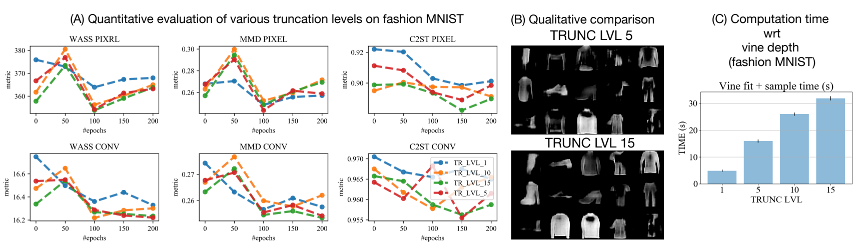
Reviewer #1

- *Introduction.* We completely agree and will further emphasize that copulas repurpose a tool learning data representations into a full-fledged generative model. Copulas allow to easily (and at a comparatively small computational cost) turn any AE into a generative model with performances that compare favorably to state-of-the-art methods. We will rewrite the intro and highlight this key point, it currently appears only in lines 68 - 75 and in Appendix F.
- *Gaussian copula.* We apologize for the confusion. In (2), $\mathcal{N}_{(\Phi^{-1}(u_j),\Phi^{-1}(v_j)),\Sigma}$ is a bivariate Gaussian distribution with mean $(\Phi^{-1}(u_j), \Phi^{-1}(v_j))$ where $\{u_j, v_j\}_{j=1}^{n}$ are the observations. (2) defines a kernel estimator of the copula density and not a Gaussian copula. We only present Gaussian copulas in Figures 3 and 4 to show how the Gaussianity assumption results in worse synthetic samples compared to the nonparametric copulas. We will clarify this in the text.
- *Typos, restructuration, and clarifying captions.* Thank you for the comments, we agree and will correct as suggested.
- *Conclusion.* An executive summary of the empirical results is indeed lacking in the conclusion and will be added.

Reviewer #2

- *Contributions.* We agree that the paper is fairly practical rather than theoretical, but simple and simplistic should not be confused. Given the general interest in generative modeling, we feel like a method allowing to repurpose AEs into generative models at a small computational cost is worthy in itself.
- *Presentation of concepts/models/algorithms.* We will extend each topic in the supplementary to make the paper as self-contained as possible. But note that the pseudo-observations are already described lines 96-97 and in Figure 2, no model selection of copulas is required since only nonparametric pairs are used, sequential estimation is described over 11 lines while referencing to the rich literature on the topic, and nonparametric estimation takes about half a page. Basic concepts such as Sklar's theorem/copula definition take only 3 lines + 4 for the density (will cut 1/2), and AEs take 11 lines before switching to generative modeling (hard to cut). Vine copulas have generated thousands of papers in the last decade and, given the space constraint, we restricted ourselves to two pages (1/4 of the paper).
- *Copula selection.* As mentioned, no selection is required since (2) (i.e., nonparametric copulas) is used for every pair.
- *Complexity.* We will add to the paper that complexity $\approx O(n \times dim \times trunc\_lvl)$ for estimation/sampling algorithms, both involving a double loop over dimension/trunc level with an internal step scaling linearly with the sample size.
- *Quality of the samples and truncation.* We will add an extended analysis. See the figure below for preliminary results suggesting that deeper vines (i.e., longer computation times) improves the quality of the generated samples. Note also the linear scaling of computation time with truncation level.
- *Continuity/differentiability.* We will add the needed assumptions for the asymptotic properties of (2) as in [19].
- *Advantages over VAEs and GANs.* Due to space constraints, the analysis detailed the comments of lines 68-75 was moved to Appendix F. We will add it back (see a similar comment from Rev#1). Regarding complexity, our claim is simply that VCAEs are easier to train. We will also describe better the results from the different metrics.
- *Typos and mistakes.* Thank you for spotting them, we agree, will correct all, and proofread better. Page 3, Appendix A.2 should be Figure 2. D_e is indeed a subset, but the other two are "elements of". Figures 2/3, yes and yes.



Reviewer #3

- *Conclusiveness of results.* This paper is a first-attempt at an alternative approach to seamlessly construct generative models by combining vines and AEs, and we chose arguably the three most common datasets and two best known competitors to illustrate our method's potential. The aim was neither an extensive empirical analysis, nor it was to prove definitive superiority against all state-of-the-art methods. In any case, Fashion MNIST will be added following the preliminary analysis from the figure above.
- *Formula 2.* Agreed, it also have been confusing to Rev#1, we could write something like $\mathcal{N}_{\mu,\Sigma}$ where $\mu = \ldots$
- *Fig 3 contours.* We should have written the contours are presented in the kernel space, i.e. after transformation to Gaussian margins (lines 172-174) and circular because X_1 and X_2 are independent.
- *Intuition for sharper images.* Two potential explanations will be added. First, blurriness in VAEs comes from the independence and Gaussianity assumptions for the latent features, but we do not assume this. Second, adding depth (trees) to the vine structure results in more dependencies/details captured, and hence hence sharper images.
- *Fig 4.* Sorry for the mix-up, the middle and right panels correspond respectively to samples for MNIST and SVNH.
- *Fig 6 for MNIST and CelebA.* The experiments were not finished by the deadline but will be added to the final version. Preliminary results indicate that the conclusions will be similar as for SVNH.
- *Fig 7 and proofreading.* Thanks for noticing about Fig 7, and proofreading was also asked by Rev#2, our apologies.
- *Memorizing in Fig 7.* Since a vine is estimated on the latent features, memorization is rather the AE's issue. Because of the struggle in the community over how to tackle memorization evaluation (see e.g.Theis et al. , 2016, Borji 2018), we thought it to be out of the score of this work, but we will mention it.