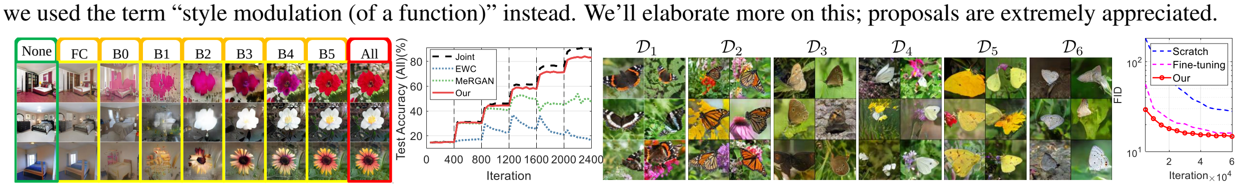
We thank all reviewers for the constructive comments. Below we first address common concerns and then respond to each reviewer to address the rest. The paper will be revised correspondingly (*e.g.,* to correct typos/errors, to move important technical details to the main paper, to revise the abstract and the broader impact carefully to better reveal the delivered information). Code will be released.

On our novelties over BSA [57] and AdaFM [95]. We consider our main novelties/contributions as (i) we reveal that one can modulate the "style" of a GAN model to form perceptually-distant but realistic targeted generation, which essentially implies GANs capture a certain universal structure to images (see L139-L141); the realistic generation from that style modulation also reveals another orthogonal dimensionality for transfer learning, which potentially outperforms/complements the commonly used finetuning (see L172-L181); (ii) we leverage (i) to deliver the GAN memory for lifelong learning, which has growing generative power yet with no forgetting; (iii) we generalize our GAN memory with compression techniques, and to *conditional* GAN. In addition to the these novelties/contributions, and generalizing FiLM/AdaFM to mFiLM/mAdaFM (see Eqs. (4)-(5)), we also empirically reveal/analyze the role each component of mFiLM/mAdaFM plays (see Sec. 4.2). None of these contributions have been made before or in [57,95] (see L84-L90 for the differences between our GAN memory and [57,95]).

On well-behaved source GAN models. By "well-behaved" we mean the shape within kernels (see L175; shared between source and target) is well trained. Empirically, this requirement can be *readily satisfied* if the source model (i) is pretrained on a (moderately) large dataset (*e.g.,* CelebA; often a dense dataset is preferred [83]) and (ii) it's sufficiently trained and shows relatively high generation quality. That means many pretrained GAN models can be "well-behaved," including the adopted GP-GAN pretrained on CelebA. One can of course expect better performance if a better source model (pretrained on a large-scale dense and diverse dataset) is used.

On the robustness to the source model. Our method is deemed considerably robust to the (pretrained) source model: (i) we modulated a different source GAN model (pretrained on LSUN Bedrooms) to form the target generation on Flowers; the resulting performance (FID=15.0) is comparable to that shown in the paper (FID=14.8 with CelebA as source); similar property about FC$\rightarrow$B6 (now B5 due to the architecture change [49]) is also observed, as shown in Fig. A; (ii) the experiments of the paper have verified that, with our style modulation process, various target domains consistently benefit from the same source model (with a better performance than finetuning), in spite of their perceptual distances to the source model; (iii) both (i) and (ii) further confirm our insights (L139-L141), *i.e.,* GANs seem to capture an underlying universal structure to images (shape within kernels (Line 175)), which may manifest as different content/semantics when modulated with different "styles;" from another perspective, (i) and (ii) also imply that universal structure may be widely carried in various "well-behaved" source models. Therefore, we believe our method and the properties in Sec. 4.2 could generalize well on different source models. We'll add more demonstrations/discussions to support our statements.

On "style". We consider our mFiLM/mAdaFM as style-transfer techniques, because they are motivated from and mathematically similar to those techniques, *i.e.,* to manipulate means and standard derivations. But the "style" here (mean/standard-derivation of kernels) is indeed different from or generalizes over style-transfer literature (see L125-L129); by "style" we mean the style *of a function* (*e.g.,* a generator, a discriminator, and potentially a classifier). For a generator, its "style" may manifest as the content of generation, which however may not fit a discriminator/classifier. To pay our respect to and distinguish from style-transfer literature, we used the term "style modulation (of a function)" instead. We'll elaborate more on this; proposals are extremely appreciated.



A. This pretrained model only has FC$\rightarrow$B5 B C D

Reviewer #1: Please see our responses above on common concerns. We'll add the suggested upperbound (labeled as "Joint" in Fig. B) and move important technical details to the main paper. We've actually considered both diverse and related/similar target tasks in Secs. 5.1 and 5.3, respectively. In Sec. 5.3, we considered 6 sequential tasks on butterfly images (one category per task; see L298-L300; Fig. C shows the generated samples). Thus, our method is believed robust to both diverse and related (image) tasks.

Reviewer #2: On one hand, our GAN memory has moderate requirements for and is considerably robust to the source model, thanks to its style modulation process (see our responses above); on the other hand, many powerful pretrained GANs have been released [95] and the valuable information therein (often benefiting downstream tasks greatly [83,95]) is one motivation for our method. In lifelong-learning settings, a growing model capacity might be necessary; when compared with MeRGAN (see L71-L78), our GAN memory works much better (see Fig. 6) with good properties (see Footnote 4); further considering its compression potential, we believe our GAN memory may serve as a practical/realistic generative replay for lifelong-learning problems. Please see our responses "On our novelties over BSA [57] and AdaFM [95]." We'll revise Table 1 following your suggestions.

Reviewer #6: Usually, our method is expected to outperform learning from scratch, because (i) our method shows better training efficiency and performance than finetuning (see Fig. 1(b)) and (ii) by referring to [83] and the transfer learning literature, finetuning from pretrained models often outperforms scratch on efficiency and performance. We empirically verified that by running scratch on Flowers (see Fig. D). To learn a GAN on (rigorous) streaming datasets (one image per task) is extremely challenging. Our streaming setting is actually a practical work-around, *e.g.,* by leveraging a physical memory buffer to form a stream of datasets with clear task boundaries. We'll add discussions and remove misleading terms. We'll enrich discussion of related work with more citations/discussions. The samples from MeRGAN are shown in the bottom two rows of Fig. 5(right); please zoom in for details.

Reviewer #9: Please see our responses "On our novelties over BSA [57] and AdaFM [95]" and "On well-behaved source GAN models." The question about $\hat{\mathbf{W}}$ is a little confusing; the notation $\odot$ denoting the Hadamard product might be misunderstood as a convolution; Eq. (3) shows how $\hat{\mathbf{W}}$ is calculated, after which $\hat{\mathbf{W}}$ is then convolved with input feature maps.