Supplementary Material

Method	IS (10 splits) (higher=better)	FID (lower=better)	Inception Accuracy (higher=better)	MS-SSIM (lower=better)	LPIPS (higher=better)
1. ImageNet-50 (real)	6.49 ± 0.40	N/A	0.90	0.43 ± 0.04	0.70 ± 0.08
2. BigGAN	6.03 ± 0.76	24.34	0.87	0.46 ± 0.05	0.61 ± 0.09
3. BigGAN + AM	6.85 ± 0.58	24.93	0.80	0.44 ± 0.03	0.64 ± 0.08
4. Noise-S	6.53 ± 0.86	28.75	0.82	0.46 ± 0.05	0.61 ± 0.09
5. Noise-L	7.67 ± 0.95	84.61	0.36	0.46 ± 0.05	0.49 ± 0.04
6. AM-S					
a. Best LPIPS trial	7.33 ± 0.73	40.82	0.72	0.44 ± 0.05	0.64 ± 0.08
b. Average	7.03 ± 0.71	38.39	0.74	0.44 ± 0.05	0.63 ± 0.08
7. AM-L					
a. Best LPIPS trial	7.49 ± 0.81	47.25	0.64	0.44 ± 0.04	0.65 ± 0.08
b. Average	7.22 ± 0.79	46.86	0.68	0.44 ± 0.05	0.63 ± 0.08
8. AM-D-S					
a. Best LPIPS trial	7.62 ± 0.90	45.61	0.66	0.44 ± 0.04	0.65 ± 0.08
b. Average	7.32 ± 0.80	43.78	0.68	0.44 ± 0.05	0.64 ± 0.08
9. AM-D-L					
a. Best LPIPS trial	7.58 ± 0.84	50.94	0.64	0.44 ± 0.04	0.65 ± 0.08
b. Average	7.43 ± 0.85	52.68	0.61	0.44 ± 0.05	0.64 ± 0.08

Table S1: We compared Activation Maximization (AM) samples with the BigGAN samples and the real ImageNet-50 images on two diversity metrics (MS-SSIM and LPIPS) and three realism metrics, Inception Score (IS), Fréchet Inception Distance (FID), and Inception Accuracy (IA). ImageNet-50 is a subset of ImageNet that contains 50 classes where BigGAN samples exhibit limited diversity (see Sec. 2.2). For each AM method, we ran 50 classes × 5 trials and reported here (a) the trial with the best LPIPS score and (b) the average across 5 runs. In MS-SSIM and LPIPS, all AM trials consistently produced more diverse samples than the BigGAN samples. However, FID and IA scores indicated that AM samples are worse in realism compared to the original BigGAN samples. See Fig. 7 for some graphical plots of this table.

S1 Explicitly encouraging diversity yielded worse sample realism

We found that in ∼2% of the AM-S and AM-L trials, the optimization converged at a class embedding that yields similar images for different random latent vectors. Here, we try to improve the sample diversity further by incorporating a specific regularization term into the AM formulation (as described in Sec. 2.1).

Experiments In the preliminary experiments, we tested encouraging diversity in the (1) image space; (2) `conv5` feature space; and (3) softmax outputs of AlexNet. We observed that the pixel-wise regularizer can improve the diversity of background colors (Fig. S1) and tends to increase the image contrast upon a

high λ multiplier (Fig. S1c). In contrast, the impact of the conv5 diversity regularizer is less noticeable (Fig. S2). Encouraging diversity in the softmax output distribution can yield novel scenes e.g. growing more flowers in monarch butterfly images (Fig. S3c).

While each level of diversity has its own benefits for specific applications, here, we chose to perform more tests with the softmax diversity to encourage samples to be more diverse *semantically*. That is, we re-ran the AM-S and AM-L experiments with an additional softmax diversity term (Eq. 3) and a coefficient $\lambda = 2$ (see Fig. S3). We call these two AM methods with the diversity term AM-D-S and AM-D-L.

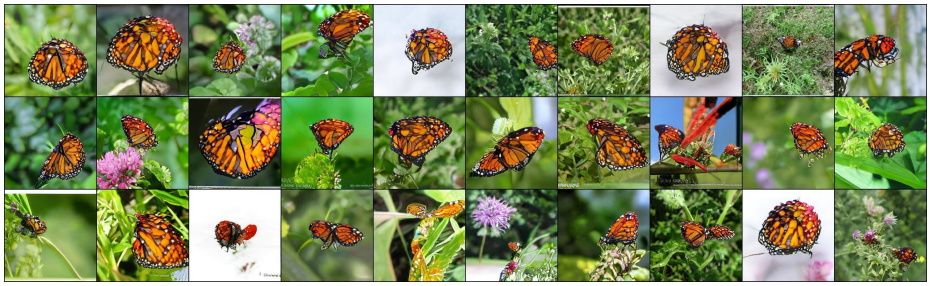
Results We found that the addition of the regularizer did not improve the diversity substantially but lowered the sample quality (Fig. 7b AM-S vs. AM-D-S and AM-L vs. AM-D-L). Similarly, the IA scores of the AM-D methods were consistently lower than those of the original AM methods (Table S1).

Method	IS (10 splits) (higher=better)	FID (lower=better)	Inception Accuracy (higher=better)	MS-SSIM (lower=better)	LPIPS (higher=better)
1. ImageNet-30 (Real)	4.18 ± 0.61	n/a	0.92	0.42 ± 0.04	0.70 ± 0.08
2. BigGAN	3.71 ± 0.74	31.36	0.91	0.45 ± 0.05	0.61 ± 0.09
3. AM-L Random					
a. AlexNet	5.06 ± 0.97	46.85	0.71	0.43 ± 0.04	0.66 ± 0.08
b. Inception-v3	4.29 ± 0.56	31.62	0.87	0.44 ± 0.04	0.65 ± 0.08
c. ResNet-50	5.36 ± 0.75	47.23	0.70	0.44 ± 0.04	0.68 ± 0.09
d. Robust ResNet-50	4.59 ± 0.69	43.65	0.76	0.43 ± 0.05	0.63 ± 0.08
4. AM-D-S					
a. AlexNet	5.31 ± 0.60	48.74	0.69	0.43 ± 0.04	0.66 ± 0.08
b. Inception-v3	4.23 ± 0.51	30.24	0.88	0.44 ± 0.04	0.65 ± 0.08
c. ResNet-50	5.78 ± 1.00	52.01	0.66	0.43 ± 0.04	0.68 ± 0.08
d. Robust ResNet-50	4.51 ± 0.79	41.74	0.78	0.44 ± 0.04	0.63 ± 0.09

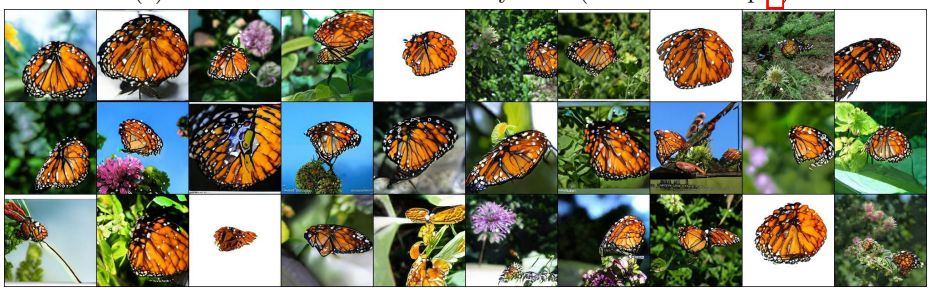
Table S2: A comparison of four different classifiers (a–d) across two preliminary AM settings across 30 random classes from the ImageNet-50 low-diversity dataset (see Sec. 2.2). The ImageNet-30 statistics here were computed from 30,000 images = 30 classes × 1000 images. Similarly, for BigGAN (Row 2) and AM-L and AM-D-S methods (Row 3–4), we generated 1000 256 × 256 samples per class. We computed the statistics for each initialization method from 5 trials, each with a different random seed. With AM-L (Sec. 3.1), we maximized the log probabilities and used a large learning rate of 0.1. With AM-D-S (Sec. 3.5), we maximized both the log probabilities and a softmax diversity regularization term, and used a small learning rate of 0.01. In sum, across both settings, AM consistently obtained the highest FID and Inception Accuracy (IA) scores with the Inception-v3 classifier (b). That is, it is possible to maximize the FID and IA scores when using Inception-v3 as the classifier in the AM formulation. However, qualitatively, we did not find the AM samples with Inception-v3 to be substantially different from the others.

Method	IS (10 splits) (higher=better)	FID (lower=better)	ResNet-18 Accuracy (higher=better)	MS-SSIM (lower=better)	LPIPS (higher=better)
1. Places-50 (real)	12.17 ± 1.01	N/A	0.57	0.42 ± 0.04	0.70 ± 0.06
2. BigGAN	8.19 ± 0.9	53.15	0.17	0.42 ± 0.05	0.66 ± 0.07
3. AM-L with Mean Initialization					
Trial 1	8.32 ± 0.89	42.38	0.51	0.43 ± 0.05	0.64 ± 0.07
Trial 2	8.39 ± 0.83	44.11	0.48	0.43 ± 0.05	0.64 ± 0.07
Trial 3	8.45 ± 0.84	42.98	0.46	0.43 ± 0.05	0.65 ± 0.07
Trial 4	7.03 ± 0.71	38.39	0.49	0.43 ± 0.05	0.64 ± 0.07
Trial 5	7.03 ± 0.71	38.39	0.49	0.43 ± 0.04	0.65 ± 0.07
Average	7.03 ± 0.51	41.25	0.49	0.43 ± 0.05	0.65 ± 0.07
4. AM-L with Top-5 Initialization					
Trial 1	8.60 ± 0.88	46.92	0.47	0.43 ± 0.05	0.65 ± 0.07
Trial 2	8.45 ± 0.81	41.09	0.52	0.43 ± 0.05	0.65 ± 0.07
Trial 3	8.13 ± 0.71	40.35	0.48	0.43 ± 0.05	0.65 ± 0.07
Trial 4	8.20 ± 0.79	43.56	0.47	0.43 ± 0.05	0.65 ± 0.07
Trial 5	8.37 ± 0.75	39.49	0.50	0.43 ± 0.05	0.65 ± 0.07
Average	8.35 ± 0.79	42.28	0.49	0.43 ± 0.05	0.65 ± 0.07

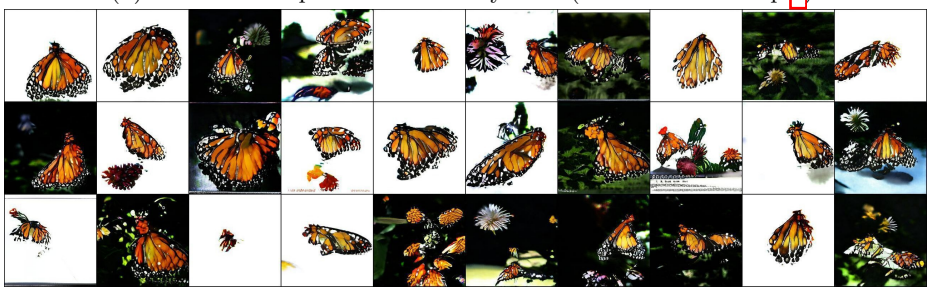
Table S3: A comparison of Places-50, BigGAN and AM images. We randomly chose 50 classes in Places365 (i.e. Places-50) to be the evaluation dataset for the experiments in Sec. 3.2. The Places-50 statistics here were computed from 50,000 images = 50 classes × 1000 images that were randomly selected from the training set of Places365. For BigGAN (Sec. 3.2), we chose the class embedding whose 10 random samples yielded the highest accuracy score for each target Places-50 class and generated 1000 samples per class. With AM-L mean initialization and AM-L top-5 initialization (Sec. 3.2), we maximized the log probabilities and used a large learning rate of 0.1. We found that samples from AM (Row 3-4) are of similar diversity but better quality than BigGAN samples.



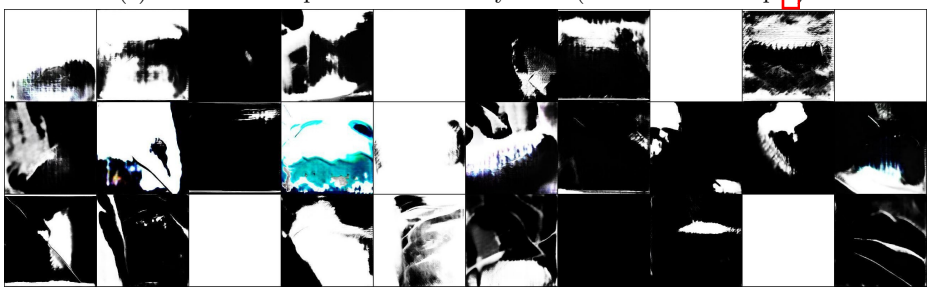
(a) AM alone without the diversity term (i.e. $\lambda = 0$ in Eq. 3).



(b) AM with the pixel-wise diversity term (i.e. $\lambda = 0.01$ in Eq. 3).

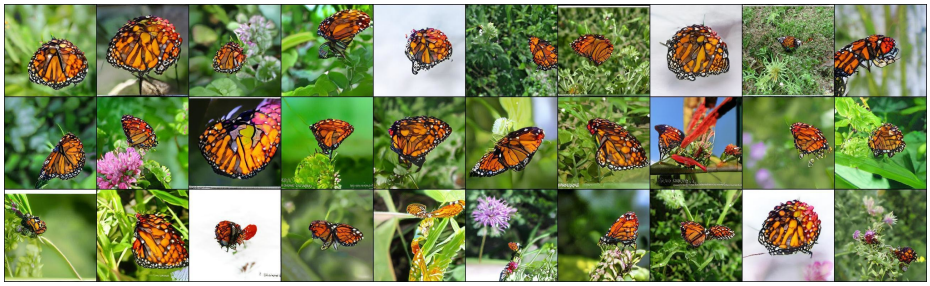


(c) AM with the pixel-wise diversity term (i.e. $\lambda = 0.1$ in Eq. 3).

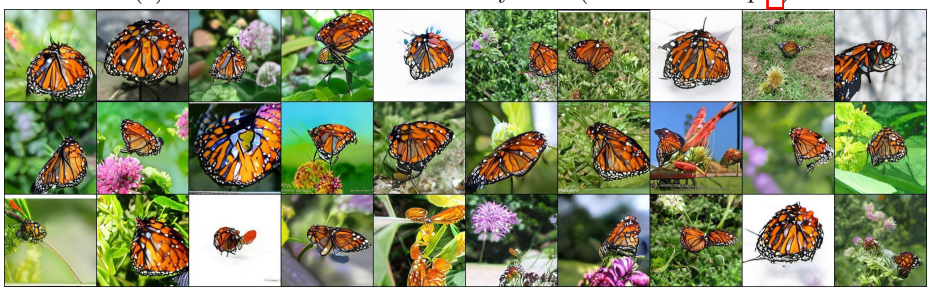


(d) AM with the pixel-wise diversity term (i.e. $\lambda = 1.0$ in Eq. 3).

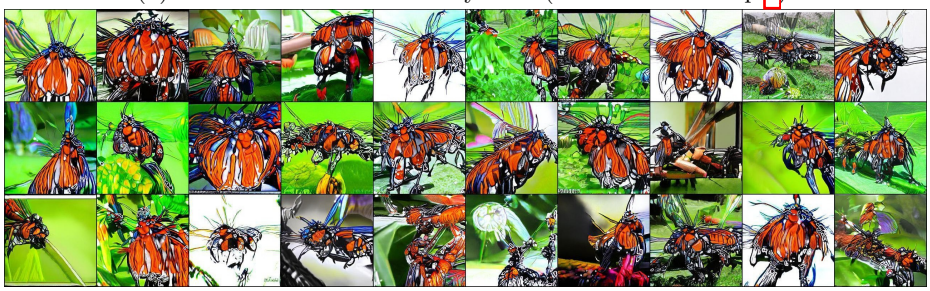
Fig. S1: The monarch butterfly class (323) samples generated by Activation Maximization (AM) methods when increasing the multiplier λ of a pixel-wise diversity regularization term in Eq. 3.



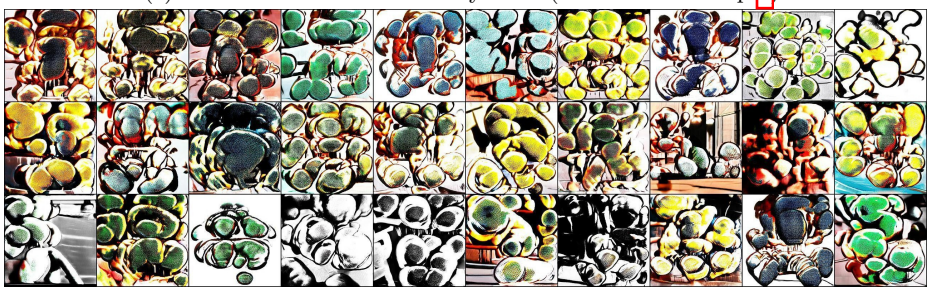
(a) AM alone without the diversity term (i.e. $\lambda = 0$ in Eq. 3).



(b) AM with a feature diversity term (i.e. $\lambda = 0.01$ in Eq. 3).

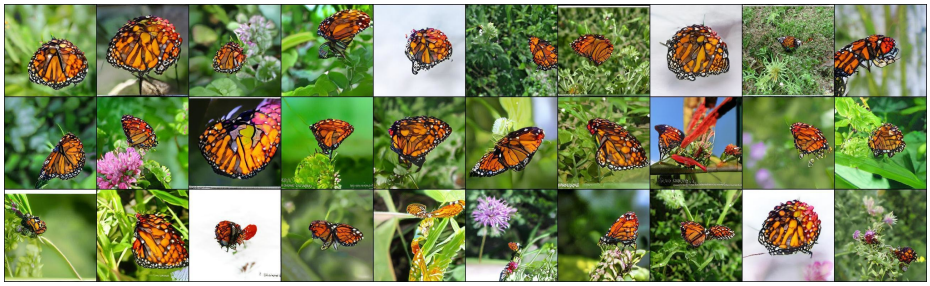


(c) AM with a feature diversity term (i.e. $\lambda = 0.1$ in Eq. 3).

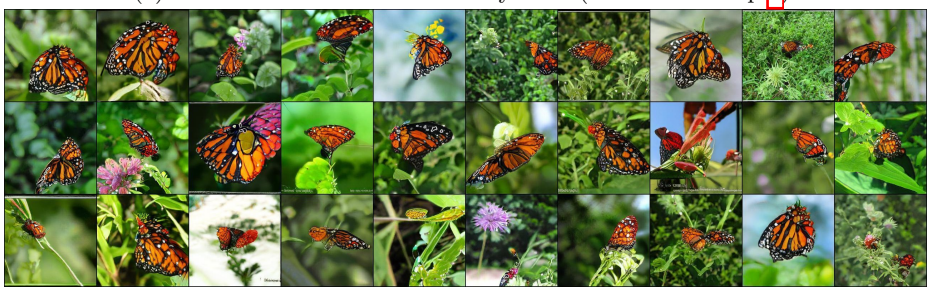


(d) AM with a feature diversity term (i.e. $\lambda = 1.0$ in Eq. 3).

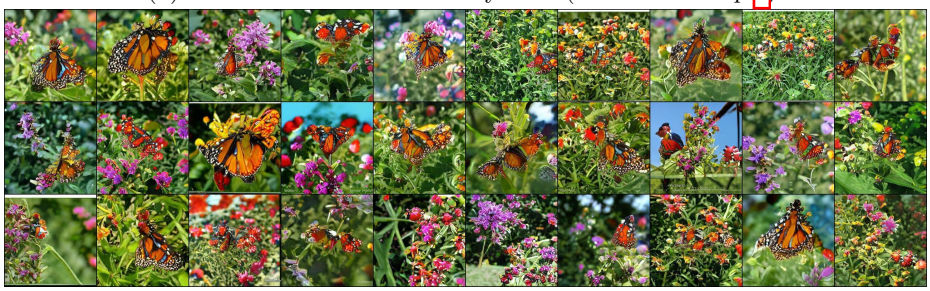
Fig. S2: The monarch butterfly class (323) samples generated by Activation Maximization (AM) methods when increasing the multiplier λ of a conv5 feature diversity regularization term in Eq. 3.



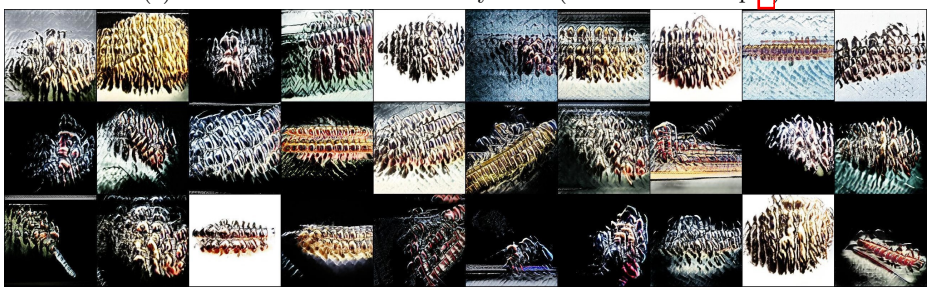
(a) AM alone without the diversity term (i.e. $\lambda = 0$ in Eq. 3).



(b) AM with a softmax diversity term (i.e. $\lambda = 2$ in Eq. 3).



(c) AM with a softmax diversity term (i.e. $\lambda = 10$ in Eq. 3).



(d) AM with a softmax diversity term (i.e. $\lambda = 100$ in Eq. 3).

Fig. S3: The monarch butterfly class (323) samples generated by Activation Maximization (AM) methods when increasing the multiplier λ of a softmax probability diversity regularization term in Eq. 3.

(a) BigGAN samples generated with the original daisy class embedding (no noise).

(b) BigGAN samples generated with the daisy class embedding $\boldsymbol{c}' = \boldsymbol{c} + \epsilon$ where noise $\epsilon \sim \mathcal{N}(0, 0.1)$.

(c) BigGAN samples generated with the daisy class embedding $\boldsymbol{c}' = \boldsymbol{c} + \epsilon$ where noise $\epsilon \sim \mathcal{N}(0, 0.3)$.

(d) BigGAN samples generated with the daisy class embedding $\boldsymbol{c}' = \boldsymbol{c} + \epsilon$ where noise $\epsilon \sim \mathcal{N}(0, 0.5)$.

Fig. S4: BigGAN samples when increasing the amount of noise added to the original daisy class embedding vector. That is, four panels (a–d) are generated using the same set of 30 latent vectors $\{\boldsymbol{z}^i\}_{30}$ but with a different class embedding $\boldsymbol{c}'$.

(A) ImageNet images (B) BigGAN samples [1]

(a) Samples from the window screen class (904).

(b) Samples from the manhole cover class (640).

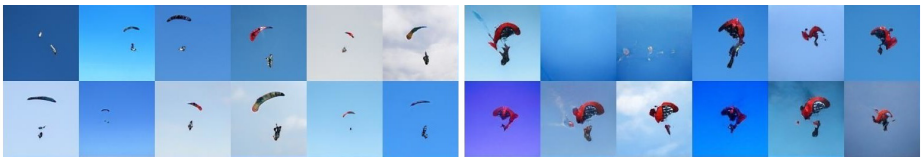
(c) Samples from the greenhouse class (580).

(d) Samples from the cardoon class (946).

Fig. S5: Example mode-collapse classes from the ImageNet-50 subset where BigGAN samples (right) exhibit substantially lower diversity compared to the real data (left).

(a) ImageNet samples from the parachute class.

(b) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 138k snapshot.



(c) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 140k snapshot.

(d) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 142k snapshot.

(e) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 144k snapshot.

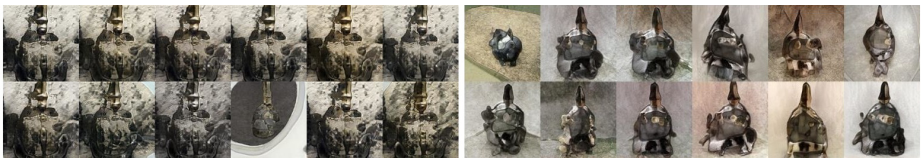
(f) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 146k snapshot.

Fig. S6: Applying our AM method to 5 different 128 × 128 BigGAN training snapshots (b–f) yielded samples (right) that qualitatively are more diverse and recognizable to be from the parachute class compared to the original BigGAN samples (left). While the original BigGAN samples are almost showing only the blue sky (d–f), AM samples show large and colorful parachutes.

(a) ImageNet samples from the pickelhaube class.

(b) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 138k snapshot.

(c) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 140k snapshot.



(d) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 142k snapshot.

(e) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 144k snapshot.

(f) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 146k snapshot.

Fig. S7: The same figure as Fig. S6 but for the pickelhaube class (715).

(a) ImageNet samples from the digital clock class.

(b) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 138k snapshot.

(c) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 140k snapshot.

(d) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 142k snapshot.

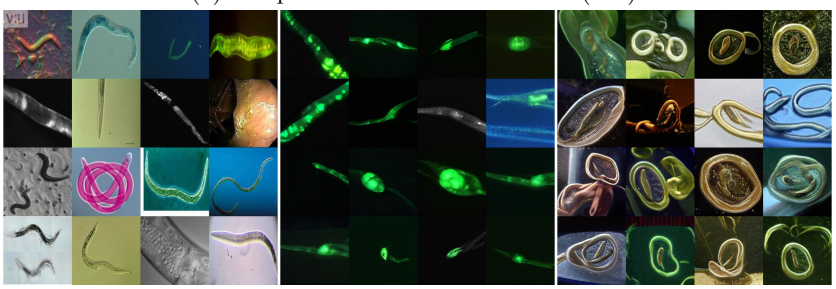
(e) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 144k snapshot.

(f) BigGAN samples (left) and AM samples (right), both generated using the BigGAN 146k snapshot.

Fig. S8: The same figure as Fig. S6 but for the digital clock class (530).

(A) ImageNet (B) BigGAN [1] (C) AM (ours)

(a) Samples from the flatworm class (110).



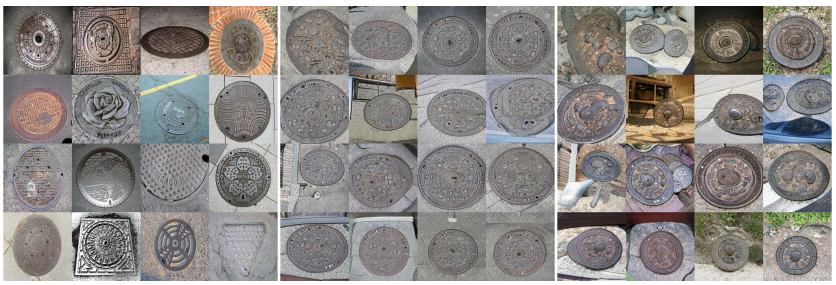
(b) Samples from the nematode class (111).

(c) Samples from the brass class (458).

(d) Samples from the greenhouse class (580).

Fig. S9: A comparison between the 256 × 256 samples from the ImageNet training set (A), the original BigGAN model (B), and our AM method (C) for four ImageNet-50 low-diversity classes. AM samples (C) are of similar quality but higher diversity than the original BigGAN samples (B). See https://drive.google.com/drive/folders/14qiLdaslnxfsCMnlBa4n1iEO1EUUYUjQ?usp=sharing for the high-resolution version of this figure.

(A) ImageNet (B) BigGAN [1] (C) AM (ours)



(a) Samples from the manhole cover class (640).

(b) Samples from the spider web class (815).

(c) Samples from the window screen class (904).

(d) Samples from the cardoon class (946).

Fig. S10: A comparison between the 256 × 256 samples from the ImageNet training set (A), the original BigGAN model (B), and our AM method (C) for four ImageNet-50 low-diversity classes. AM samples (C) are of similar quality but higher diversity than the original BigGAN samples (B). See https://drive.google.com/drive/folders/14qiLdaslnxfsCMnlBa4n1iEO1EUUYUjQ?usp=sharing for the high-resolution version of this figure.

(A) ImageNet (B) BigGAN [1] (C) AM (ours)

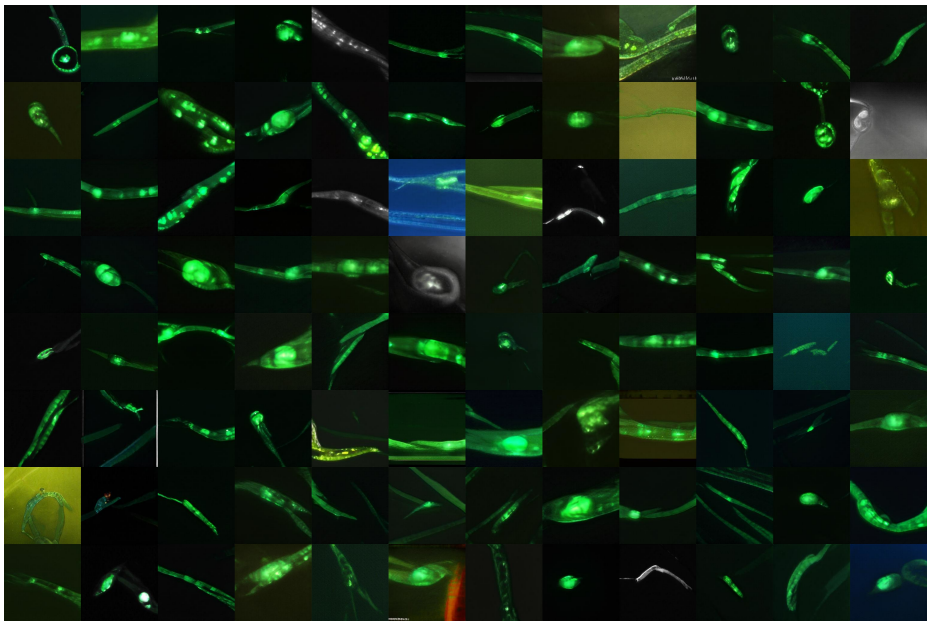
(a) Samples from the pineapple class (953).

(b) Samples from the custard apple class (956).

(c) Samples from the carbonara class (959).

(d) Samples from the pizza class (963).

Fig. S11: A comparison between the 256 × 256 samples from the ImageNet training set (A), the original BigGAN model (B), and our AM method (C) for four ImageNet-50 low-diversity classes. AM samples (C) are of similar quality but higher diversity than the original BigGAN samples (B). See https://drive.google.com/drive/folders/14qiLdaslnxfsCMnlBa4n1iEO1EUUYUjQ?usp=sharing for the high-resolution version of this figure.



(a) Samples from BigGAN.

(b) Samples from AM.

Fig. S12: A comparison between the 256×256 samples from the original BigGAN model (a), and our AM method (b) for the nematode class (111). AM samples (b) are of similar quality but higher diversity than the original BigGAN samples (a).

(a) Samples from BigGAN.

(b) Samples from AM.

Fig. S13: A comparison between the 256×256 samples from the original BigGAN model (a), and our AM method (b) for the brass class (458). AM samples (b) are of similar quality but higher diversity than the original BigGAN samples (a).

(a) Samples from BigGAN.

(b) Samples from AM.

Fig. S14: A comparison between the 256×256 samples from the original BigGAN model (a), and our AM method (b) for the greenhouse class (580). AM samples (b) are of similar quality but higher diversity than the original BigGAN samples (a).

(a) Samples from BigGAN.

(b) Samples from AM.

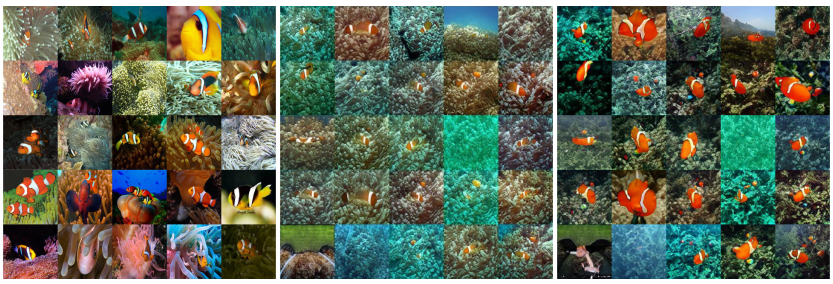
Fig. S15: A comparison between the 256×256 samples from the original BigGAN model (a), and our AM method (b) for the window screen class (904). AM samples (b) are both of higher quality and higher diversity than the original BigGAN samples (a).

(a) Samples from BigGAN.

(b) Samples from AM.

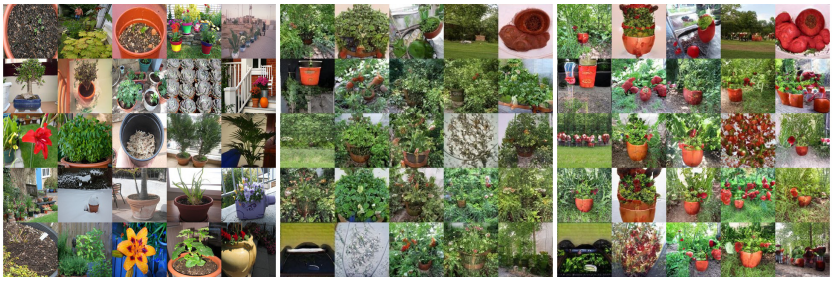
Fig. S16: A comparison between the 256×256 samples from the original BigGAN model (a), and our AM method (b) for the daisy class (985). AM samples (b) are of similar quality but higher diversity than the original BigGAN samples (a).

(A) ImageNet (B) BigGAN [1] (C) AM (ours)



(a) Samples from the anemone fish class (393).

(b) Samples from the odometer class (685).

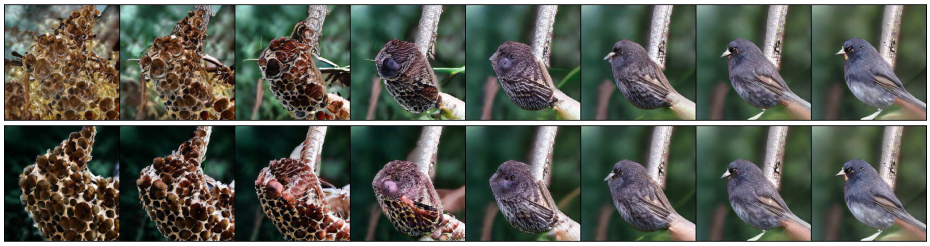


(c) Samples from the flowerpot class (738).

(d) Samples from the consomme class (925).

Fig. S17: A comparison between the 128×128 samples from the ImageNet training set (A), the original BigGAN model (B), and our AM method (C) for four ImageNet-50 low-diversity classes. AM samples (C) are of similar quality but higher diversity than the original BigGAN samples (B).

(a) Interpolation in the embedding space between seaurchin (leftmost) and German shepherd (rightmost).



(b) Interpolation in the embedding space between honeycomb (leftmost) and junco bird (rightmost).

(c) Interpolation in the embedding space between hot pot (leftmost) and cheeseburger (rightmost).

Fig. S18: The interpolation samples between $\boldsymbol{c}$ class-embedding pairs with latent vectors $\boldsymbol{z}$ held constant. In each panel, the top row shows the interpolation between two original 256×256 BigGAN embeddings while the bottom row shows the interpolation between an embedding found by AM (leftmost) and the original BigGAN embedding (right). In sum, the interpolation samples with the AM embeddings (bottom panels) appear to be similarly plausible as the original BigGAN interpolation samples (top panels).

(a) Interpolation in the embedding space between window screen (leftmost) and water tower (rightmost).

(b) Interpolation in the embedding space between espresso (leftmost) and pop bottle (rightmost).

(c) Interpolation in the embedding space between agaric (leftmost) and bolete (rightmost).

Fig. S19: The interpolation samples between $\boldsymbol{c}$ class-embedding pairs (from related ImageNet classes e.g. agaric and bolete are both mushrooms) with latent vectors $\boldsymbol{z}$ held constant. In each panel, the top row shows the interpolation between two original 256×256 BigGAN embeddings while the bottom row shows the interpolation between an embedding found by AM (leftmost) and the original BigGAN embedding (right). In sum, the interpolation samples with the AM embeddings (bottom panels) appear to be similarly plausible as the original BigGAN interpolation samples (top panels).

(a) Interpolation in the latent space between two $\boldsymbol{z}$ vectors with the same greenhouse class embedding.

(b) Interpolation in the latent space between two $\boldsymbol{z}$ vectors with the same window screen class embedding.

(c) Interpolation in the latent space between two $\boldsymbol{z}$ vectors with the same espresso class embedding.

(d) Interpolation in the latent space between two $\boldsymbol{z}$ vectors with the same daisy flower class embedding.

Fig. S20: The interpolation samples between $\boldsymbol{z}$ latent-vector pairs with the same class embeddings. The $\boldsymbol{z}$-interpolation samples with the AM embeddings (bottom panels) appear to be similarly plausible as the original BigGAN interpolation samples (top panels). For the window screen class (b), AM recovered the human-unrecognizable BigGAN samples into a plausible interpolation between two scenes of windows.

Fig. S21: A comparison between the 256×256 samples from the Places365 training set (A), the BigGAN samples generated for the ImageNet class whose 10 random samples were given the highest accuracy for the target class in Places365 (B), and our AM samples (C). AM samples (C) are of similar diversity but better quality than the original BigGAN samples (B). See `https://drive.google.com/drive/folders/1L-1ULPfOf_5-98I7emYW86OPDu3Fjxnx?usp=sharing` for a high-resolution version of this figure.

(A) Places365 (B) BigGAN on ImageNet (C) AM (ours)

Fig. S22: The same figure as Fig. S21 but for four different classes. While the ImageNet axolotl class samples were given the highest accuracy (bottom panel), they are qualitatively more different from the real jacuzzi images compared to the AM samples which shows the bathtubs. See https://drive.google.com/drive/folders/1L-1ULPfOf_5-98I7emYW86OPDu3Fjxnx?usp=sharing for a high-resolution version of this figure.

Fig. S23: The same figure as Fig. S21 but for four different classes. In the bottom panel, while the BigGAN samples are dock images that contain mostly ships whereas AM samples show more bridges that resemble the real pier samples in Places365. See https://drive.google.com/drive/folders/1L-1ULPfOf_5-98I7emYW86OPDu3Fjxnx?usp=sharing for a high-resolution version of this figure.

(A) Places365 (B) BigGAN on ImageNet (C) AM (ours)

Fig. S24: The same figure as Fig. S21 but for four different classes. For the baseball stadium, the top-1 ImageNet class is scoreboard (B), an object commonly found in stadiums. However, the AM samples are more similar to the images from Places365, which often do not contain scoreboards (A vs. C). See https://drive.google.com/drive/folders/1L-1ULPfOf_5-98I7emYW86OPDu3Fjxnx?usp=sharing for a high-resolution version of this figure.

(a) alcove (b) beach house (c) boathouse

(d) jacuzzi indoor (e) pagoda (f) pier

(g) railroad track (h) baseball stadium) (i) synagogue outdoor

Fig. S25: For each class, we find 2 class embeddings by using AM and generate a set of images by using the same $\boldsymbol{z}$. The samples from each class have different style corresponding to different class embeddings.