

DIVERSITY AUGMENTED CONDITIONAL GENERATIVE ADVERSARIAL NETWORK FOR ENHANCED MULTIMODAL IMAGE-TO-IMAGE TRANSLATION - SUPPLEMENTARY MATERIAL

Anonymous authors

Paper under double-blind review

A ANALYSIS OF THE DIVERSITY AUGMENTED REGULARIZATION

DivAugGAN introduces a novel regularization for cGANs to promote local sensitivity. It directly augments the diversity of generated samples, i.e., $G(\mathbf{x}, \mathbf{z}_r)$ with the latent style code \mathbf{z}_r . A large norm of the first-order derivative ensures the sensitive responses to style codes, and a moderate norm of the second-order derivative encourages weak decay of the sensitivity. Motivated from this point, we formulate the DivAugGAN regularizer as:

$$\mathcal{L}_{da} = \max_G \mathbb{E}_{\mathbf{z}_r} \left\{ \lambda_1 \left\| \frac{\partial G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{z}} \right\| - \lambda_2 \left\| \frac{\partial^2 G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{z}^2} \right\| \right\}. \quad (1)$$

Since a conditional generator $G(\mathbf{x}, \mathbf{z}) : \mathbb{R}^\ell \times \mathbb{R}^k \rightarrow \mathbb{R}^d$ is a multivariate function with high-dimensional outputs, it is very difficult to optimize the norms of its derivatives explicitly. As illustrated in Figure 1, we consider to approximating these norms with the average norms of the corresponding *directional derivatives* along any random direction $\mathbf{v} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_k)$, which can be further estimated with *finite difference methods* from data pairs $(G(\mathbf{x}, \mathbf{z}_r), \mathbf{x}, \mathbf{z}_r, \Delta \mathbf{z})$.

Directional Derivatives. For a multivariate function $g(\mathbf{x}) : \mathbb{R}^\ell \rightarrow \mathbb{R}$ and a directional vector $\mathbf{v} \in \mathbb{R}^\ell$, its first-order and second-order directional derivatives are closely related to the corresponding derivatives, i.e., $\|\mathbf{v}\|_2 \frac{dg(\mathbf{x})}{d\mathbf{v}} = \mathbf{v}^T \frac{dg(\mathbf{x})}{d\mathbf{x}}$, $\|\mathbf{v}\|_2^2 \frac{d^2g(\mathbf{x})}{d\mathbf{v}^2} = \mathbf{v}^T \frac{d^2g(\mathbf{x})}{d\mathbf{x}^2} \mathbf{v}$. By replacing the norms in Eq. 1 with average norms of the corresponding *directional derivatives* along a random direction $\mathbf{v} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_k)$, we define the transformed regularization as:

$$\mathcal{L}_{da} = \max_G \mathbb{E}_{\mathbf{z}_r} \left\{ \lambda_1 \mathbb{E}_{\mathbf{v}} \left[\left\| \frac{\partial G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}} \right\| \right] - \lambda_2 \mathbb{E}_{\mathbf{v}} \left[\left\| \frac{\partial^2 G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}^2} \right\| \right] \right\}, \quad (2)$$

where $\frac{\partial G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}}$ and $\frac{\partial^2 G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}^2}$ refer to the first-order partial directional derivative and the second-order partial directional derivative of $G(\mathbf{x}, \mathbf{z})$ to \mathbf{z} , respectively.

When ℓ_1 norm is employed, we can prove the following proportional expression holds between the first-order norms:

$$\mathbb{E}_{\mathbf{v} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_k)} \left[\left\| \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial \mathbf{v}} \right\| \right] \propto \left\| \frac{\partial G(\mathbf{x}, \mathbf{z})}{\partial \mathbf{z}} \right\|. \quad (3)$$

This proportional expression implies that average norms of directional derivatives can function as a surrogate for norms of derivatives, especially in formulating regularization losses.

Finite Differences of Directional Derivatives. Finite difference methods provide a simple and useful technique to approximate directional derivatives of multivariate functions. Taking the first-order derivative as an example, forward, backward and central difference approximations of $\frac{\partial G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}}$ can be defined as:

$$\begin{aligned} \left\| \frac{\partial G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}} \right\| &= \frac{\|G(\mathbf{x}, \mathbf{z}_r + \beta \Delta \mathbf{z}) - G(\mathbf{x}, \mathbf{z}_r)\|}{\|\beta \Delta \mathbf{z}\|_2} + \mathcal{O}(\beta \|\Delta \mathbf{z}\|_2), \\ \left\| \frac{\partial G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}} \right\| &= \frac{\|G(\mathbf{x}, \mathbf{z}_r) - G(\mathbf{x}, \mathbf{z}_r - \alpha \Delta \mathbf{z})\|}{\|\alpha \Delta \mathbf{z}\|_2} + \mathcal{O}(\alpha \|\Delta \mathbf{z}\|_2), \\ \left\| \frac{\partial G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}} \right\| &= \frac{\|G(\mathbf{x}, \mathbf{z}_r + \beta \Delta \mathbf{z}) - G(\mathbf{x}, \mathbf{z}_r - \alpha \Delta \mathbf{z})\|}{\|(\alpha + \beta) \Delta \mathbf{z}\|_2} + \mathcal{O}((\alpha + \beta)^2 \|\Delta \mathbf{z}\|_2^2), \end{aligned} \quad (4)$$

where the principal parts on the right hand side are named as DR terms of the diversity augmented regularization.

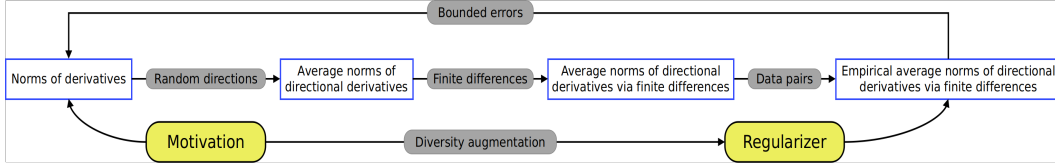


Figure 1: Theoretical derivation of the diversity augmented regularizer in DivAugGAN.

For the second-order derivative $\frac{\partial^2 G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}^2}$, it holds that

$$\begin{aligned}
 \left\| \frac{\partial^2 G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}^2} \right\| &= \frac{\|G(\mathbf{x}, \mathbf{z}_r + \beta \Delta \mathbf{z}) - 2G(\mathbf{x}, \mathbf{z}_r) + G(\mathbf{x}, \mathbf{z}_r - \beta \Delta \mathbf{z})\|}{\|\beta \Delta \mathbf{z}\|_2^2} + \mathcal{O}(\beta^3 \|\Delta \mathbf{z}\|_2^3), \\
 \left\| \frac{\partial^2 G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}^2} \right\| &= \frac{\|G(\mathbf{x}, \mathbf{z}_r + 2\beta \Delta \mathbf{z}) - 2G(\mathbf{x}, \mathbf{z}_r + \beta \Delta \mathbf{z}) + G(\mathbf{x}, \mathbf{z}_r)\|}{\|\beta \Delta \mathbf{z}\|_2^2} + \mathcal{O}(\beta^2 \|\Delta \mathbf{z}\|_2^2), \\
 \left\| \frac{\partial^2 G(\mathbf{x}, \mathbf{z}_r)}{\partial \mathbf{v}^2} \right\| &= \frac{\|G(\mathbf{x}, \mathbf{z}_r) - 2G(\mathbf{x}, \mathbf{z}_r - \alpha \Delta \mathbf{z}) + G(\mathbf{x}, \mathbf{z}_r - 2\alpha \Delta \mathbf{z})\|}{\|\alpha \Delta \mathbf{z}\|_2^2} + \mathcal{O}(\alpha^2 \|\Delta \mathbf{z}\|_2^2),
 \end{aligned} \tag{5}$$

where the principal parts on the right hand side are named as RVC terms of the diversity augmented regularization.

B APPLICATION TO UNCONDITIONAL GAN

Apart from conditional GAN, we further explore the regularization effect of DivAugGAN on unconditional GAN with a mixture of twenty-five 2D Gaussians arranged as a 5×5 square matrix. In order to reduce regularization from network structures, we simply set the generator and the discriminator as a three-layer MLP and a two-layer MLP, respectively, without any normalization. We collect 5000 real samples as training data and visualize 2500 generated samples. For all regularization methods, the best regularization coefficient is chosen from $\{1, 10, 50, 100, 200\}$. As illustrated in Figure 2, the mode-collapse problem can be effectively alleviated by applying our DivAugGAN regularizer to encourage the generator to efficiently explore the data space. DivAugGAN regularizer successfully enables the generator to capture much more modes, when compared to vanilla GAN, DSGAN and MSGAN setting.

C RELATED WORKS

GANs Goodfellow et al. (2014) have demonstrate remarkable effectiveness in various computer vision and graphics tasks, e.g., image/video synthesis, image/video translation, domain adaptation and data augmentation. cGANs, built upon GANs, takes additional information as extra conditional inputs, and can be applied to various applications. However, cGANs are often observed to suffer mode collapse problems Arjovsky & Bottou (2017), resulting in only small subsets of output distribution are represented by the generator. In image-to-image translation task, this issue leads to a deterministic mapping from input to output distributions, and multi-modal nature of the mapping is sacrificed.

Extensive studies have been performed to resolve the commonly appeared mode collapse problem in both of standard and conditional GAN settings, such as incorporating the mini-batch statistics into the discriminator Salimans et al. (2016), employing the improved divergence metrics, objective functions and optimization processes to smooth the loss of the discriminator Arjovsky et al. (2017); Gulrajani et al. (2017); Mao et al. (2017); Odena et al. (2018); Heusel et al. (2017); Srivastava et al. (2017); Miyato et al. (2018), and introducing auxiliary networks, like multiple generators or discriminators with weight-sharing mechanism Liu & Tuzel (2016); Ghosh et al. (2018); Hoang et al. (2018); Nguyen et al. (2017); Che et al. (2016), extra encoders Dumoulin et al. (2017); Donahue et al. (2017); Larsen et al. (2016) and additional classifier Odena et al. (2017); Lin et al. (2019), etc. Hybrid model of cGAN and VAE with random injected latent codes is also presented to address

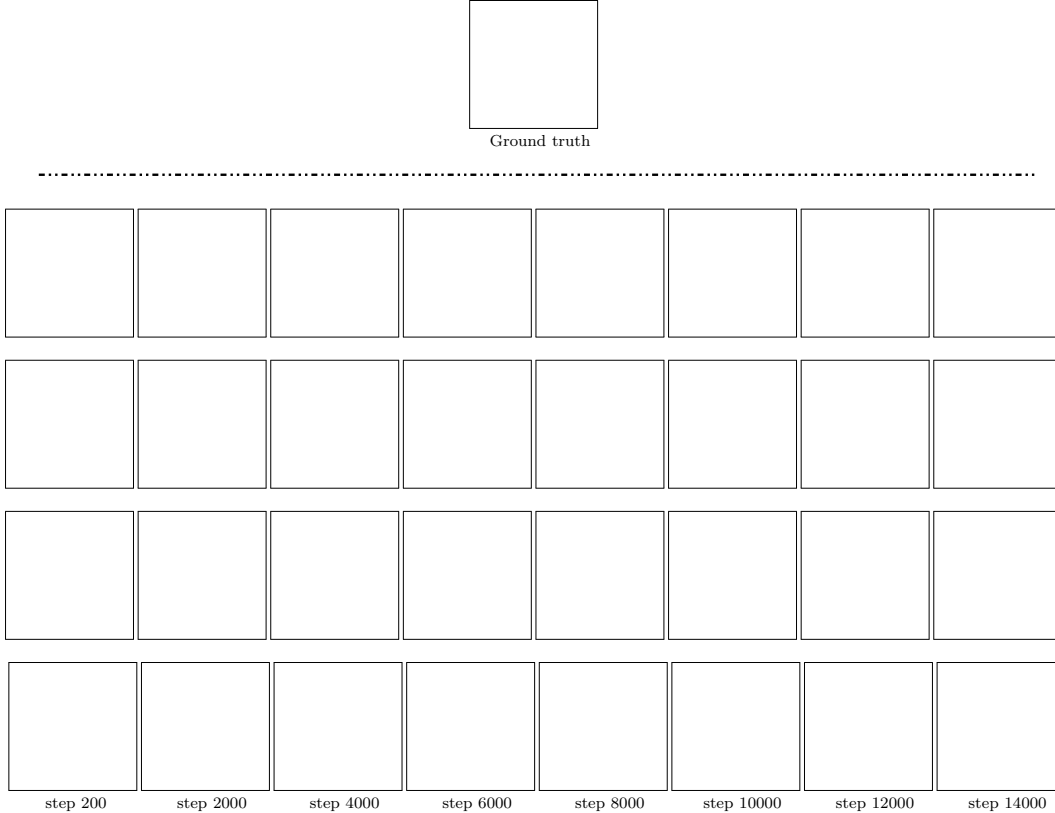


Figure 2: Density plots of ground-truth, vanilla GAN (1st row), DSGAN (2nd row), MSGAN (3rd row) and DivAugGAN (4th row). **Please zoom the image to see the details.**

mode-collapse issue in cGANs based framework for multimodal image-to-image translation task. Specifically, Zhu *et al.* design an invertible generator in BiCycleGAN with an additional encoder network for latent code reconstruction from the generated image. Similarly, domain-specific decoders are developed to interpret the latent codes for generating images with various styles in multi-modal image translation by Lee *et al.* Lee *et al.* (2018) and Huang *et al.* Huang *et al.* (2018), respectively.

Odena *et al.* propose a regularization method to clamp the generator Jacobian within a certain range Odena *et al.* (2018). Sharing a similar idea as Odena *et al.* (2018), Yang *et al.* presented DSGAN with an objective function to simply maximize the norm of the generator gradient with an optional upper-bound Yang *et al.* (2019), and Mao *et al.* proposed MSGAN with an additional mode seeking regularization term on the generator to maximize the ratio of the distance between the produced images with respect to the distance between the injected latent vectors. However, all such regularization fail to maintain the relative change coherence, as they only include the absolute distinction requirements while ignore to present relative consistency constraints, which may bring about unexpected mode override or mode fusion issues.

D TRAINING DETAILS

We train all experiments on Tesla V100 GPUs 32G using PyTorch Paszke *et al.* (2019).

Two-domain multimodal image-to-image translation. All models are trained using Adam Kingma & Ba (2015) with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ and batch size is 12. We also use a weight decay at rate of 0.0001. We train all experiments on *Dog* \rightleftharpoons *Cat* and *Summer* \rightleftharpoons *Winter* datasets for 2000 and 1200 epochs, respectively. We keep the same learning rate of 0.0001 for the first 600 epochs and linealy decay the rate to the last epoch. For training data, we resized them to 256×256 , and random cropped them to 216×216 . For testing data, we resized them to 256×256 , and center cropped them to 216×216 . The network weights are initialized with Gaussian initialization.

Multi-domain multimodal image-to-image translation. For MDMM, MDMM+MSGAN and DivAugGAN(M), the hyper-parameters keep same with two-domain image-to-image translation experiments. Except that, *AFHQ dataset* is trained for 2000 epochs, and the other three are trained for 1200 epochs. For StarGANv2 and DivAugGAN(s), we using the Adam with $\beta_1 = 0$ and $\beta_2 = 0.99$ and batch size is 8. We also use a weight decay at rate of 0.0001. We train all experiments 100K iterations. The learning rates of *Generator*, *Discriminator* and *Style encoder* are set to 10^{-4} , while that of *Mapping network* is set to 10^{-6} . For training data, we random cropped them with the scale 0.8 to 1.0, then resized to 256×256 . They are flipped horizontally with a probability of 0.5. For testing data, we just resized them to 256×256 . The network weights are initialized with Kaiming initialization He et al. (2015).

E EVALUATION DETAILS

Dog \Rightarrow Cat dataset Lee et al. (2018). There are 771 and 1264 training images in the class of *Cat* and *Dog*, respectively. The number of test images in each class is 100.

Summer \Rightarrow Winter dataset Zhu et al. (2017). This dataset contains images downloaded from Flickr with the tag *yosemite*. The training size of each class is 1232 (*Summer*) and 962 (*Winter*) and the test size of each class is 309 (*Summer*) and 238 (*Winter*)

Alps Seasons dataset Anoosheh et al. (2018). This data set is collected from images on Flickr. The images are categorized into four seasons based on the provided timestamp of when it was taken. It consists of four categories: *Spring*, *Summer*, *Fall*, and *Winter*. The training data consists of 6053 images of four seasons, while the test data consists of 400 images.

WikiArts dataset Zhu et al. (2017). This data set includes painting images of four artists *Monet*, *Van Gogh*, *Cezanne*, and *Ukiyo-e*, and another real photo data set. There are 1072, 525, 400, 562, and 6287 images, for the class of *Monet*, *Van Gogh*, *Cezanne*, *Ukiyo-e*, and *real photo*, respectively, in the training set, while there are 121, 58, 400, 263, and 751 images, for the class of *Monet*, *Van Gogh*, *Cezanne*, *Ukiyo-e*, and *real photo*, respectively, in the test set.

Image weather conditions dataset Chu et al. (2017). We only select four different weather condition images, i.e., *sunny*, *cloudy*, *snowy*, and *foggy* in this work. The training data consists of 1202, 1202, 1202, and 307 images for *sunny*, *cloudy*, *snowy*, and *foggy* weather condition, respectively, while the test set consists of 50 images for each class.

AFAQ dataset Choi et al. (2020). This data sets includes three domains of animal face images: *cat*, *dog*, and *wildlife*. For each domain, we use 4500 images as the training set, and remain 500 images as the test set. Note that in AFAQ dataset, all images are vertically and horizontally aligned to make the eyes at the center.

E.1 EVALUATION METRICS

We conduct quantitative evaluations using the following metrics.

Frechét inception distance (FID) Heusel et al. (2017). *FID* measures the distance between the distributions of the two images sets. Following Choi et al. (2020), the feature vectors from the last average pooling layer of the ImageNet Deng et al. (2009); Russakovsky et al. (2015) pretrained Inception-V3 network Szegedy et al. (2016) are used. We translate each test image from the source domain into the target domain with 10 randomly sampled latent vectors. Then, the *FID* between the translated images and training images in the target domain is calculated for each domain pairs and the average scores are reported. Lower *FID* values indicate better quality of the generated images.

Learned perceptual image patch similarity (LPIPS) Zhang et al. (2018). *LPIPS* measures the diversity of the generated images using the L1 distance between the extracted features from the pretrained AlexNet Krizhevsky et al. (2012). Following Choi et al. (2020), we generate 10 output images of a target domain using 10 randomly sampled latent vectors for each test image from the source domain. Next, the average of the pairwise distances among all 45 pairs outputs generated from the same input is calculated, and the mean LPIPS values for all test images are reported. Note that higher LPIPS scores indicate better diversity among the generated images.

Number of statistically-different bins (NDB) Richardson & Weiss (2018). *NDB* is a bin-based metric for measuring the similarity between the distribution between real generated and images. *NDB* metric is employed to measure level of the mode collapse. Following Richardson & Weiss (2018), we first execute a standard K-means algorithm with 10 random initializations to cluster the training samples into different bins ($N = 100, 200$ or 300), which represents the modes of the real data distribution. Next, each generated sample is assigned to its nearest bin. Note that each bin center is the mean of all samples assigned to the same cluster. In the third step, the bin-proportions of the training and synthesized samples are computed respectively to evaluate the difference between the generated distribution and the real data distribution. Lastly, we find *NDB* of the bin-proportion by determining the mode missing extent. Note that lower *NDB* scores indicate the generated data distribution fits the real data distribution better with more modes.

Jensen-Shannon divergence (JSD) Richardson & Weiss (2018). We also compute the *JSD* metric between the reference bins distribution and the tested model bins distribution, which is used as an alternative to the *NDB* metric.

$$J(P||Q) = \frac{1}{2}\sum P(x_i)\log\frac{P(x_i)}{M(x_i)} + \frac{1}{2}\sum Q(x_i)\log\frac{Q(x_i)}{M(x_i)}, \text{ where } M(x_i) = \frac{P(x_i)+Q(x_i)}{2}$$

E.2 REFERENCE MODELS

We compare the performance of the proposed DivAugGAN with the following reference models:

DRIT Lee et al. (2018). *DRIT* is a two domain multimodal image translation framework trained with unpaired data.

MSGAN Mao et al. (2019). This method uses a mode seeking regularization to alleviate the mode-collapse problem in conditional generation tasks. Given a conditional image I , latent vectors z_1 and z_2 , and a conditional generator G , it use the mode seeking regularization term to maximize the ratio of the distance between $G(I, z_1)$ and $G(I, z_2)$ with respect to the distance between z_1 and z_2 .

MDMM Lee et al. (2020). This method uses a single generator G and a single discriminator D to perform translation among multiple domains. Given K domains $\{K_i\}_{i=1\sim k}$ and their one-hot domain codes $\{z_i^d\}_{i=1\sim k}$, the method encodes the images onto a shared content space C , and domain-specific attribute spaces $\{A_i\}_{i=1\sim k}$. It uses the target domain code, the target domain attribute and a random content to generate a target image. The discriminator not only aims to discriminate between real images and translated images, but also performs domain classification.

StarGANv2 Choi et al. (2020). This method trains a single generator G that can generate diverse images of multiple domains. It consists of four modules, a generator, a mapping network, a style encoder and a discriminator. The mapping network works to transform a latent code into domain-specific style codes for multiple domains and the style encoder is used to extract the style code of an input image. Then, the generator uses the domain-specific style code to translate an input image to the target domain. The discriminator consists of multiple output branches, each branch responsible for a certain domain to determine whether an image is a real image of its domain or a fake image produced by G .

F ADDITIONAL RESULTS

F.1 MORE TWO-DOMAIN MULTIMODAL IMAGE-TO-IMAGE TRANSLATION RESULTS

Figures 3, 4 and 5 present more two-domain multimodal image-to-image translation results on *dog \Rightarrow cat*, *Yosemite summer \Rightarrow winter*, and *photo \Rightarrow cezanne/monet/ukiyo-e/vangogh* datasets. DivAugGAN generates much diverse outputs over DRIT and MSGAN.

F.2 MORE MULTI-DOMAIN MULTIMODAL IMAGE-TO-IMAGE TRANSLATION RESULTS

Comparison on weather condition dataset. As the quantitative experimental results exhibited in Table ??, the proposed DivAugGAN performs favorably against MDMM and StarGANv2 in all metrics for this multi-domain translation task on the *weather condition* dataset. DivAugGAN generates much diverse outputs over MDMM with superior image visual quality. For example, in *sunny \rightarrow foggy*



Figure 3: More qualitative comparisons of DivAugGAN with DRIT and MSGAN on $dog \rightleftharpoons cat$ and Yosemite $summer \rightleftharpoons winter$ for two-domain multimodal image-to-image translation tasks.



Figure 4: More qualitative comparisons of DivAugGAN with DRIT and MSGAN on *dog \rightleftharpoons cat* and Yosemite *summer \rightleftharpoons winter* for two-domain multimodal image-to-image translation tasks.

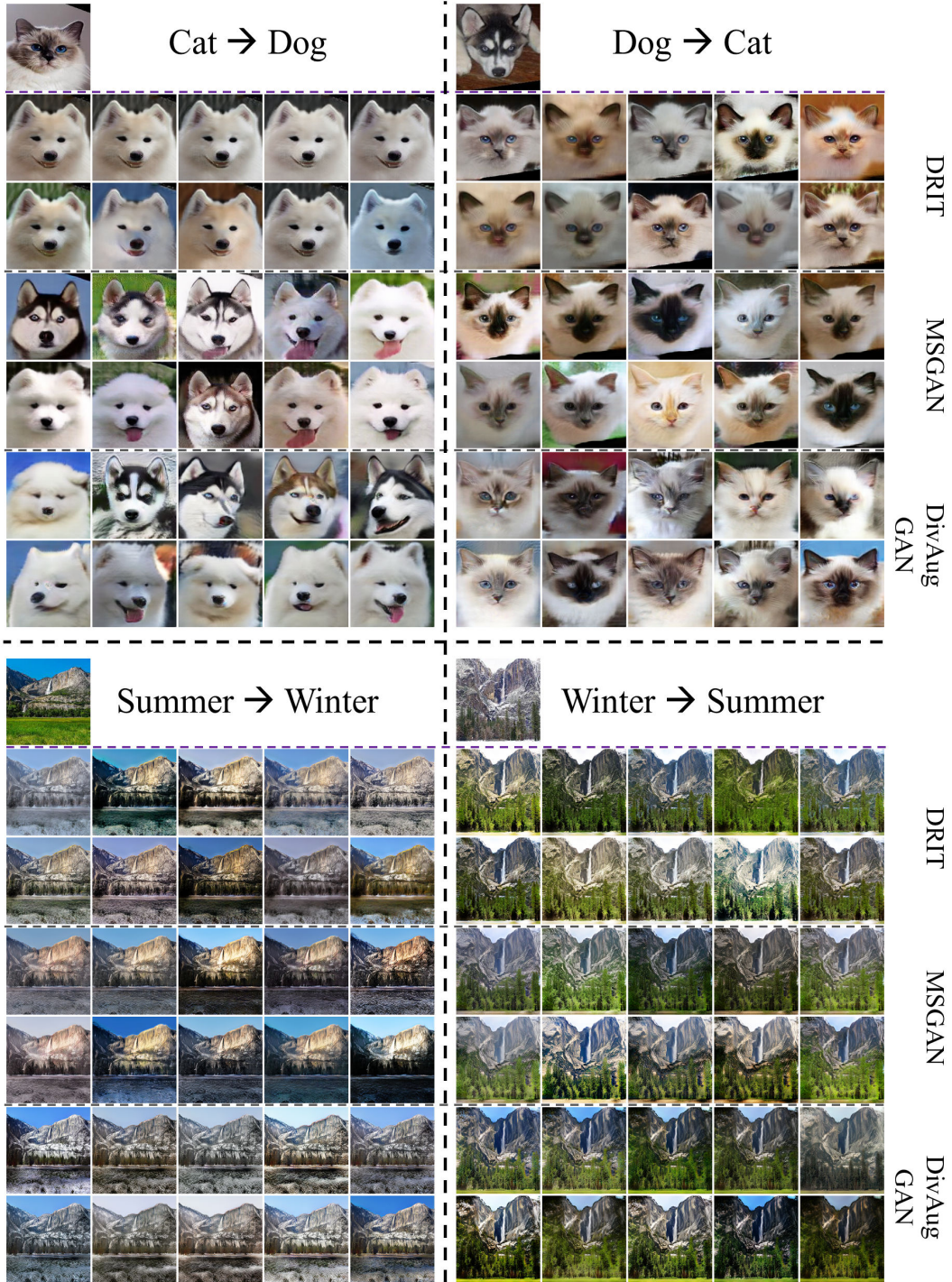


Figure 5: More qualitative comparisons of DivAugGAN with DRIT and MSGAN on *dog \rightleftharpoons cat* and Yosemite *summer \rightleftharpoons winter* for two-domain multimodal image-to-image translation tasks.

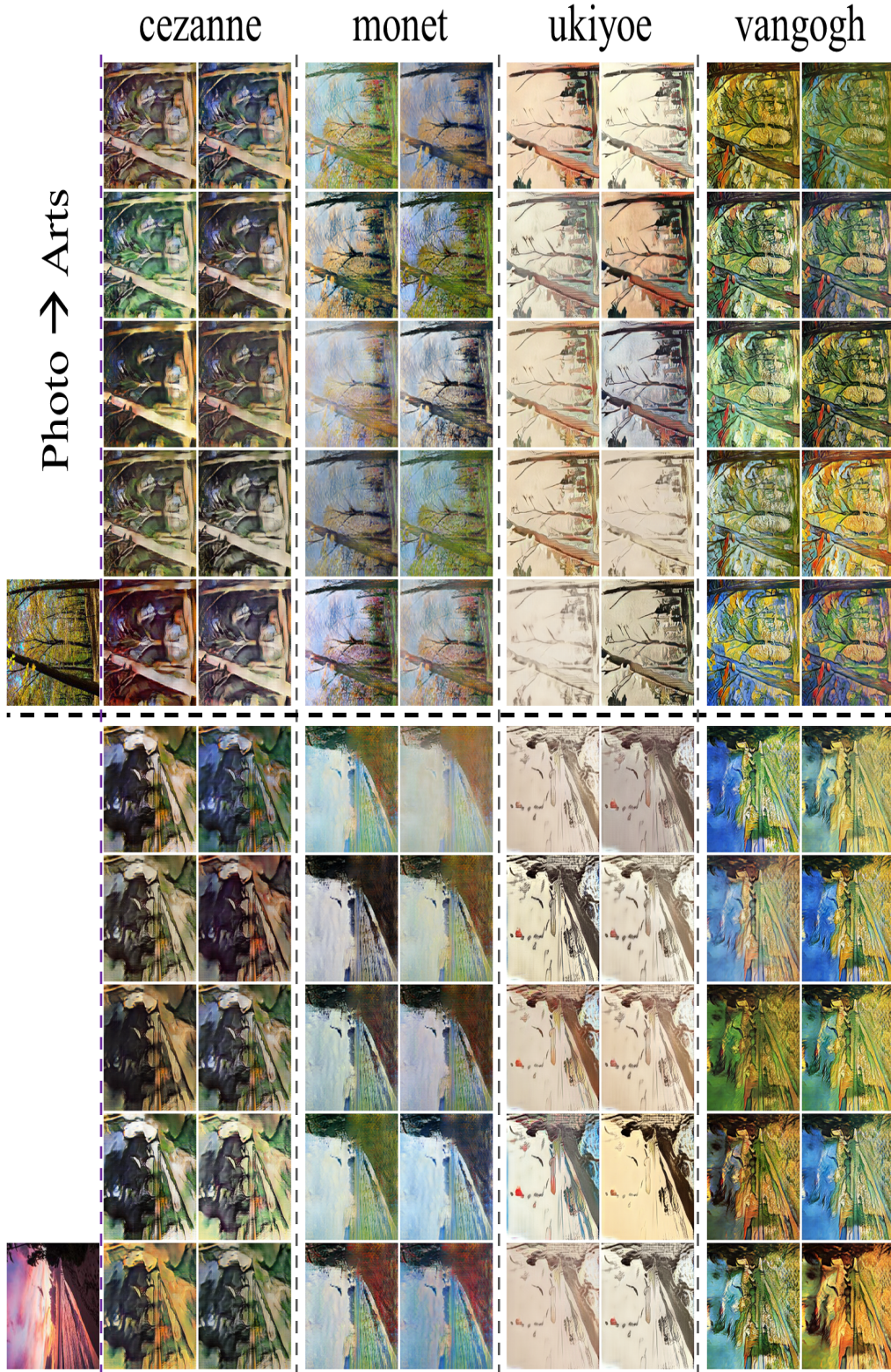


Figure 6: More qualitative results of DivAugGAN on $photo \Rightarrow \text{cezanne/monet/ukiyoe/vangogh}$ for two-domain multimodal image-to-image translation tasks. Rotate the figure 90 degrees clockwise to see.

Table 1: Quantitative comparison results of MDMM, MDMM with MSGAN regularizer and MDMM with DivAugGAN regularizer on *weather condition* dataset.

		FID ↓	LPIPS ↑	NDB ↓	JSD ↓
MDMM	Cloudy → Sunny	125.61±0.28	0.0658±0.0012	26.00±2.00	0.164±0.002
	Snow → Sunny	177.33±0.66	0.0739±0.0025	22.33±2.67	0.157±0.005
	Foggy → Sunny	135.76±0.76	0.0658±0.0035	27.00±2.00	0.206±0.004
	Sunny → Cloudy	119.36±0.25	0.0637±0.0017	17.33±0.67	0.110±0.005
	Snow → Cloudy	174.90±0.68	0.0722±0.0031	25.00±2.00	0.169±0.003
	Foggy → Cloudy	131.44±0.54	0.0629±0.0015	19.00±3.00	0.110±0.002
	Sunny → Snow	170.42±0.41	0.0683±0.0022	26.00±1.00	0.190±0.007
	Cloudy → Snow	167.92±0.30	0.0695±0.0020	28.67±2.33	0.248±0.007
	Foggy → Snow	160.11±0.32	0.0627±0.0023	28.67±0.67	0.261±0.003
	Sunny → Foggy	146.50±0.24	0.0637±0.0017	13.00±1.00	0.130±0.005
MDMM +MSGAN	Cloudy → Foggy	148.39±0.30	0.0662±0.0027	8.67±0.67	0.084±0.004
	Snow → Foggy	198.48±0.48	0.0691±0.0031	15.67±1.33	0.228±0.009
	Cloudy → Sunny	118.64±0.56	0.1102±0.0028	30.00±1.00	0.179±0.004
	Snow → Sunny	168.57±0.47	0.1315±0.0013	21.33±1.33	0.157±0.006
	Foggy → Sunny	137.41±0.54	0.1130±0.0043	25.00±3.00	0.206±0.003
	Sunny → Cloudy	118.16±0.39	0.1131±0.0023	14.00±2.00	0.108±0.004
	Snow → Cloudy	167.08±0.63	0.1206±0.0047	24.33±2.67	0.172±0.005
	Foggy → Cloudy	133.17±0.70	0.1020±0.0038	19.67±1.33	0.112±0.004
	Sunny → Snow	165.74±0.64	0.1091±0.0026	24.33±2.33	0.184±0.013
	Cloudy → Snow	165.05±0.83	0.1032±0.0027	26.33±0.67	0.247±0.010
DivAug GAN(M)	Foggy → Snow	165.46±0.22	0.1020±0.0017	30.00±1.00	0.284±0.008
	Sunny → Foggy	145.76±0.79	0.1059±0.0007	11.67±2.67	0.124±0.007
	Cloudy → Foggy	144.54±0.51	0.1053±0.0042	8.00±1.00	0.096±0.007
	Snow → Foggy	188.93±0.43	0.1141±0.0037	19.00±1.00	0.275±0.010
	Cloudy → Sunny	116.74±0.19	0.1596±0.0052	19.33±1.33	0.146±0.008
	Snow → Sunny	150.86±0.92	0.1736±0.0041	18.00±2.00	0.134±0.011
	Foggy → Sunny	126.90±0.57	0.1630±0.0012	20.33±0.67	0.193±0.010
	Sunny → Cloudy	109.01±0.70	0.1624±0.0036	15.33±3.67	0.094±0.002
	Snow → Cloudy	148.54±0.20	0.1784±0.0030	23.67±1.67	0.170±0.004
	Foggy → Cloudy	125.71±0.51	0.1595±0.0024	19.33±0.67	0.090±0.007
	Sunny → Snow	148.16±0.84	0.1570±0.0042	23.33±1.33	0.185±0.005
	Cloudy → Snow	153.84±1.37	0.1616±0.0097	26.00±2.00	0.218±0.005
	Foggy → Snow	147.19±0.92	0.1619±0.0084	29.67±1.67	0.244±0.009
	Sunny → Foggy	140.13±0.55	0.1650±0.0063	14.67±0.67	0.151±0.005
	Cloudy → Foggy	150.87±0.55	0.1699±0.0034	11.00±1.00	0.139±0.010
	Snow → Foggy	186.51±0.32	0.1776±0.0089	18.67±0.67	0.220±0.007

translation task, we achieve much higher *LPIPS* score, i.e. 0.1650 [DivAugGAN(M)] v.s. 0.1059 [MDMM+MSGAN] v.s. 0.0637 [MDMM], and lower *FID* score, i.e., 140.13 [DivAugGAN(M)] v.s. 145.76 [MDMM+MSGAN] v.s. 146.50 [MDMM]. Figures 7, 8, 9, and 10 present the complete results of the qualitative comparisons of DivAugGAN (M) with MDMM and MSGAN integrated MDMM on *image weather condition* dataset for multi-domain multimodal image-to-image translation. We show all twelve translation results, including *cloudy* → *foggy*, *cloudy* → *snow*, *cloudy* → *sunny*, *foggy* → *cloudy*, *foggy* → *snow*, *foggy* → *sunny*, *snow* → *foggy*, *snow* → *cloudy*, *snow* → *sunny*, *sunny* → *cloudy*, *sunny* → *snow*, and *sunny* → *foggy*.

Qualitative and quantitative comparisons on AFAQ data set. As the quantitative experimental results exhibited in Table 2, the proposed DivAugGAN performs favorably against MDMM and StarGANv2 in most metrics for this multi-domain multimodal image-to-image translation task. DivAugGAN achieves slightly diverse outputs over MDMM and StarGANv2. For example, in *dog* → *cat* translation, we achieve higher *LPIPS* score, i.e. 0.5141 [DivAugGAN(S)] v.s. 0.4956 [StarGANv2], 0.4433 [DivAugGAN(M)] v.s. 0.4177 [MDMM+MSGAN] v.s. 0.2901 [MDMM]. Note that StarGANv2 is specifically designed for this task. Figure 11 presents the complete results of the qualitative comparisons DivAugGAN (S) with StarGANv2 on *AFHQ* dataset for multi-domain multimodal image-to-image translation. We show all six translation results within three domains:

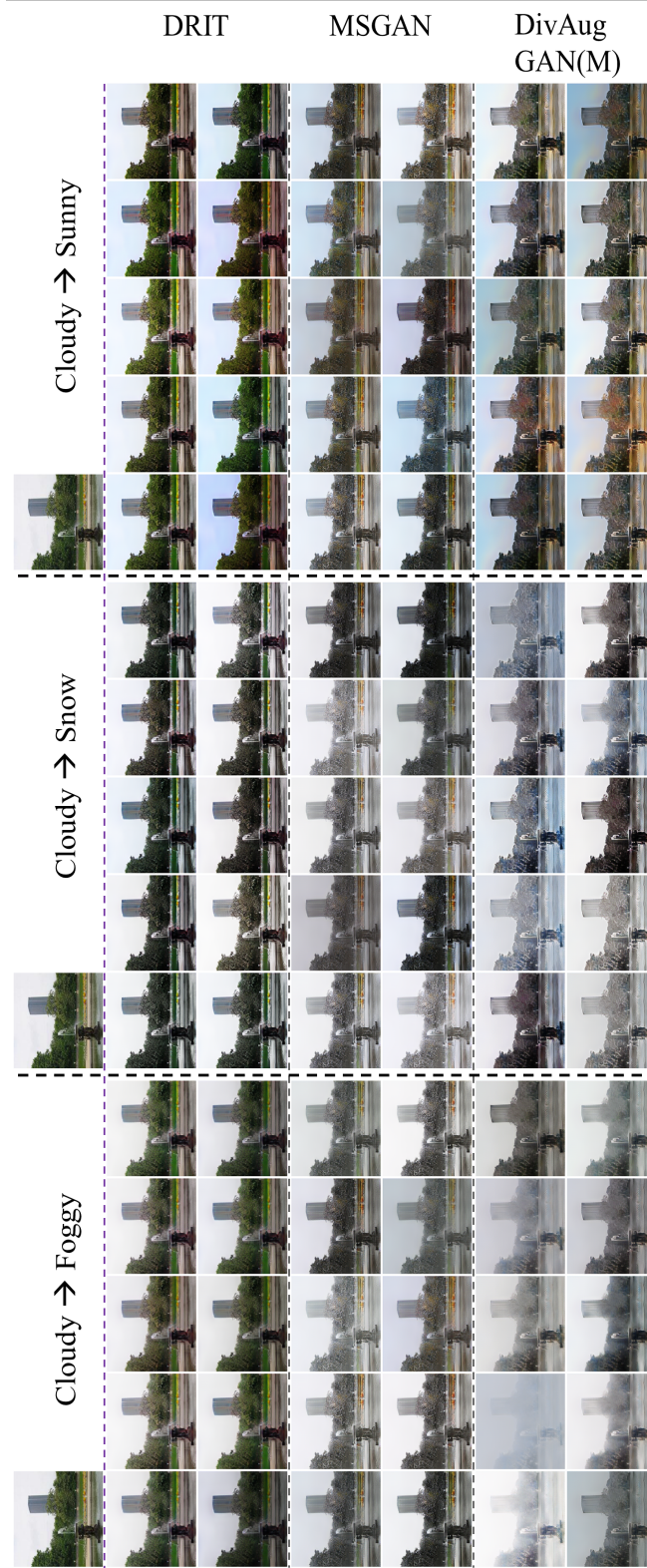


Figure 7: Qualitative comparisons of DivAugGAN (M) with MDMM and MSGAN integratd MDMM on image weather condition dataset for multi-domain multimodal image-to-image translation. Translation results of cloudy \rightarrow foggy/snow/sunny are illustrated. Rotate the figure 90 degrees clockwise to see.

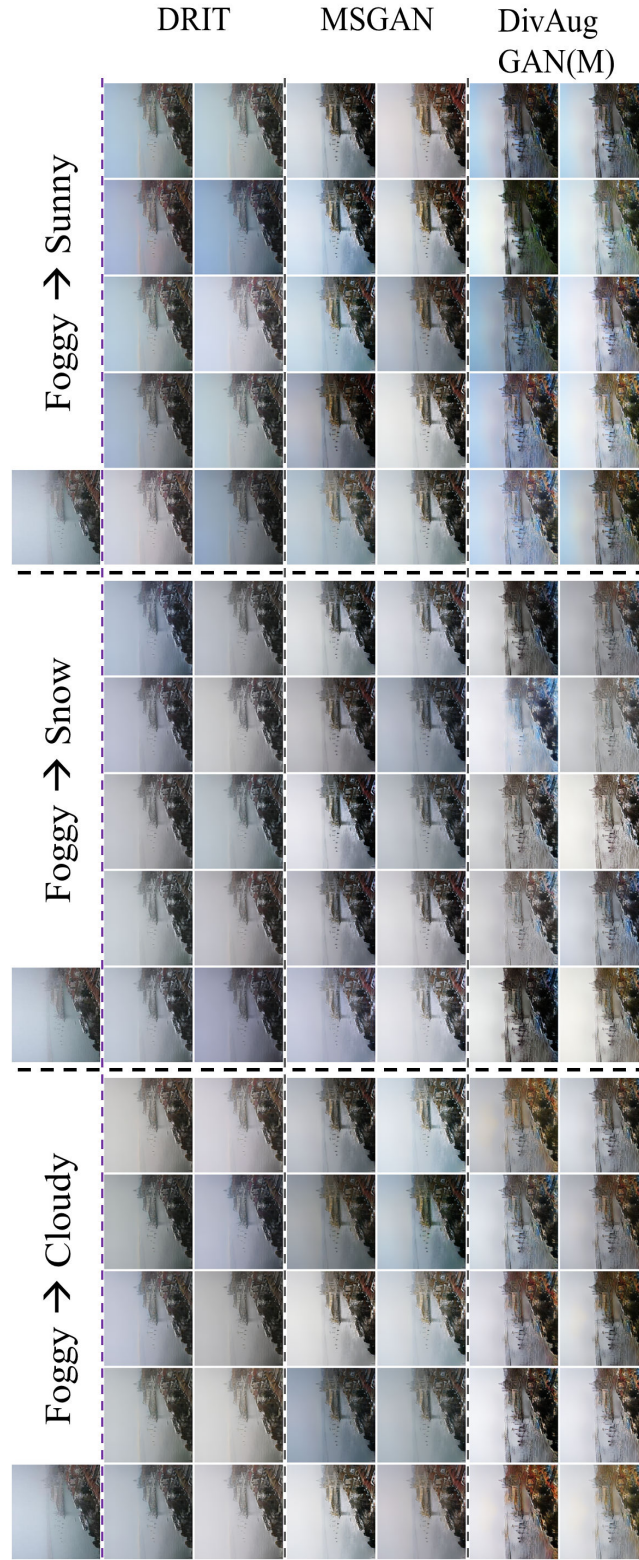


Figure 8: Qualitative comparisons of DivAugGAN (M) with MDMM and MSGAN integratd MDMM on image weather condition dataset for multi-domain multimodal image-to-image translation. Translation results of foggy \rightarrow cloudy/snow/sunny are illustrated. Rotate the figure 90 degrees clockwise to see.

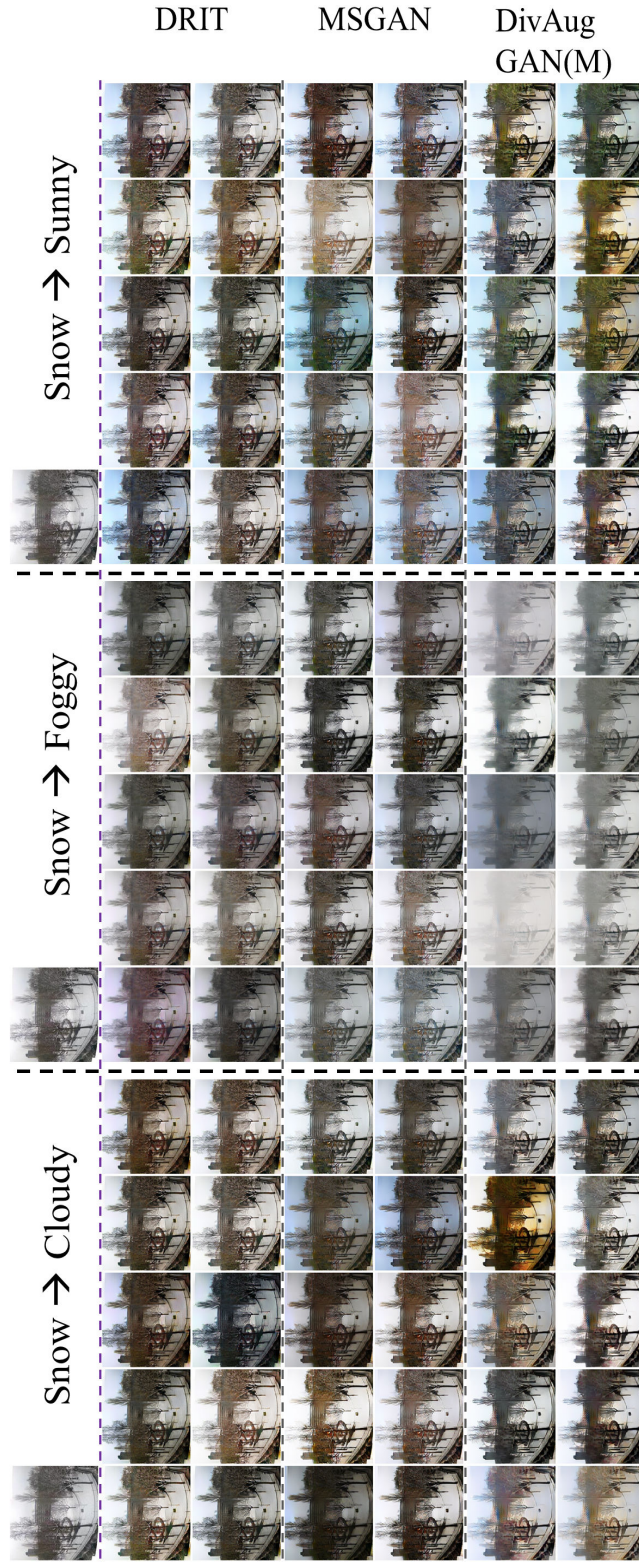


Figure 9: Qualitative comparisons of DivAugGAN (M) with MDMM and MSGAN integratd MDMM on image weather condition dataset for multi-domain multimodal image-to-image translation. Translation results of snow \rightarrow cloudy/foggy/sunny are illustrated. Rotate the figure 90 degrees clockwise to see.



Figure 10: Qualitative comparisons of DivAugGAN (M) with MDMM and MSGAN integratd MDMM on image weather condition dataset for multi-domain multimodal image-to-image translation. Translation results of sunny → cloudy/foggy/snow are illustrated. Rotate the figure 90 degrees clockwise to see.

Table 2: Quantitative comparison results of MDMM Lee et al. (2020), MDMM with MSGAN Mao et al. (2017) regularizer, StarGANv2 Choi et al. (2020), MDMM with DivAugGAN regularizer, and StarGANv2 with DivAugGAN regularizer on *AFHQ* dataset.

		FID ↓	LPIPS ↑	NDB ↓	JSD ↓
MDMM	Dog → Cat	40.61±0.42	0.2901±0.0032	38.00±2.00	0.070±0.002
	Wild → Cat	23.40±0.23	0.3001±0.0015	40.00±2.00	0.096±0.001
	Cat → Dog	66.02±0.23	0.3591±0.0015	34.67±1.67	0.064±0.001
	Wild → Dog	49.49±0.24	0.3353±0.0010	35.67±0.67	0.070±0.002
	Cat → Wild	46.31±0.07	0.3198±0.0010	34.67±0.67	0.072±0.001
	Dog → Wild	72.30±0.36	0.2544±0.0014	40.00±1.00	0.078±0.008
MDMM +MSGAN	Dog → Cat	12.22±0.18	0.4177±0.0012	36.00±2.00	0.072±0.001
	Wild → Cat	13.01±0.30	0.4050±0.0008	40.00±2.00	0.075±0.002
	Cat → Dog	31.18±0.03	0.4880±0.0005	43.67±1.67	0.095±0.004
	Wild → Dog	28.01±0.74	0.4762±0.0003	39.33±2.33	0.102±0.001
	Cat → Wild	19.24±0.26	0.4611±0.0013	42.33±1.33	0.169±0.004
	Dog → Wild	24.34±0.60	0.4642±0.0010	39.67±0.67	0.170±0.004
DivAug GAN(M)	Dog → Cat	13.85±0.20	0.4433±0.0022	39.00±1.00	0.125±0.002
	Wild → Cat	15.96±0.15	0.4378±0.0004	39.33±0.67	0.122±0.004
	Cat → Dog	50.75±0.30	0.4860±0.0012	42.00±2.00	0.110±0.001
	Wild → Dog	49.39±0.12	0.4881±0.0004	36.67±1.67	0.120±0.004
	Cat → Wild	24.65±0.85	0.4584±0.0022	37.00±2.00	0.097±0.002
	Dog → Wild	24.31±0.80	0.4642±0.0007	37.00±1.00	0.090±0.002
StarGANv2	Dog → Cat	8.59±0.25	0.4956±0.0006	21.33±1.67	0.024±0.001
	Wild → Cat	6.83±0.17	0.4876±0.0009	26.33±4.33	0.025±0.002
	Cat → Dog	39.50±0.66	0.5197±0.0006	25.67±1.67	0.026±0.003
	Wild → Dog	33.16±0.57	0.5206±0.0039	29.67±1.67	0.032±0.004
	Cat → Wild	14.24±0.10	0.4889±0.0001	41.00±1.00	0.109±0.005
	Dog → Wild	14.38±0.33	0.4921±0.0004	36.67±3.33	0.111±0.004
DivAug GAN(S)	Dog → Cat	9.01±0.17	0.5141±0.0008	40.00±1.00	0.117±0.001
	Wild → Cat	6.65±0.11	0.5114±0.0020	42.33±2.67	0.119±0.004
	Cat → Dog	36.52±0.15	0.5296±0.0006	43.67±1.33	0.130±0.008
	Wild → Dog	32.49±0.32	0.5287±0.0013	46.33±1.33	0.128±0.003
	Cat → Wild	14.05±0.13	0.4893±0.0007	30.33±0.67	0.044±0.001
	Dog → Wild	13.46±0.31	0.4879±0.0017	29.33±0.67	0.046±0.002

wild → *cat*, *cat* → *wild*, *wild* → *dog*, *dog* → *wild*, *cat* → *dog*, and *dog* → *cat*. The produced images by both of DivAugGAN(M) and DivAugGAN(S) shows superior visual quality.

Qualitative and quantitative comparisons on *alps seasonal transfer* dataset. As the quantitative experimental results presented in Table 3, the proposed DivAugGAN performs favorably against MDMM and StarGANv2 with a margins in all metrics for this multi-domain translation task. DivAugGAN generates much diverse outputs over MDMM and StarGANv2 with superior image visual quality. For example, in *winter* → *spring* translation task, we achieve much higher *LPIPS* score, i.e., 0.4312 [DivAugGAN(S)] vs 0.2966 [StarGANv2], 0.1803 [DivAugGAN(M)] v.s. 0.1352 [MDMM+MSGAN] v.s. 0.0869 [MDMM], and lower *FID/JSD* scores, i.e., 58.26/0.051 [DivAugGAN(S)] v.s. 68.84/0.073 [StarGANv2], 78.27/0.045 [DivAugGAN(M)] v.s. 82.05/0.066 [MDMM+MSGAN] v.s. 89.22/0.076 [MDMM]. Figure 12, 13, 14, and 15, present the complete results of qualitative comparison of DivAugGAN with MDMM and StarGANv2 on *alps seasonal transfer* dataset for multi-domain multimodal image-to-image translation. We show all twelve translation results with in fourth domains: *spring* → *summer*, *spring* → *autumn*, *spring* → *winter*, *summer* → *spring*, *summer* → *autumn*, *summer* → *winter*, *autumn* → *spring*, *autumn* → *summer*, *autumn* → *winter*, *winter* → *spring*, *winter* → *summer*, and *winter* → *autumn*.

More qualitative results on *WikiArts* dataset. More results of our DivAugGAN(S) are shown in Figure 16 to demonstrate its effectiveness and generality. Translation results of *photo* → *cezanne/monet/ukiyo-e/vangogh* are illustrated.

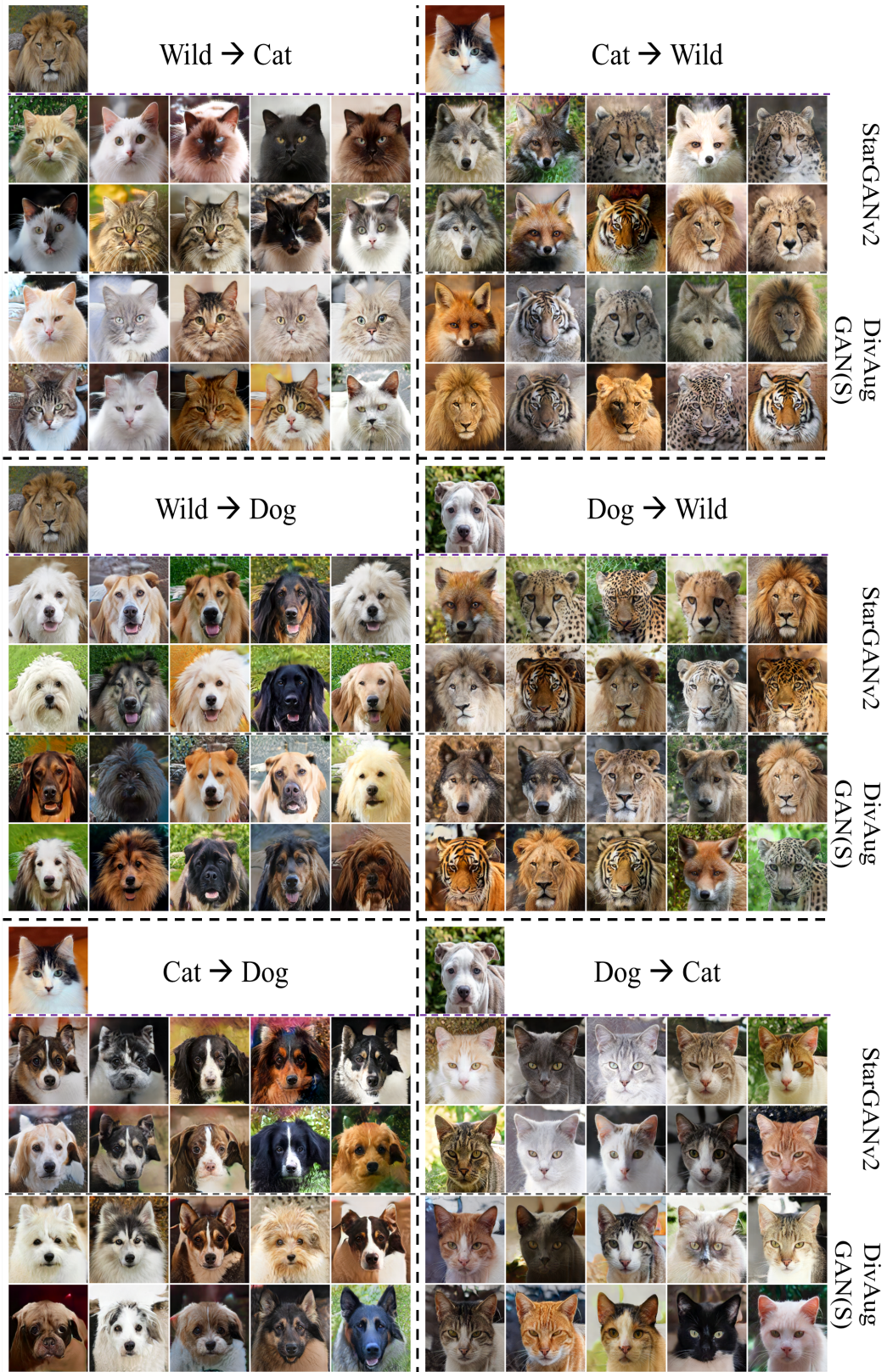


Figure 11: Qualitative comparisons of DivAugGAN (S) with StarGANv2 on AFHQ dataset for multi-domain multimodal image-to-image translation. Complete translation results of $Wild \rightleftharpoons Cat$, $Wild \rightleftharpoons Dog$, and $Dog \rightleftharpoons Cat$ are illustrated.

Table 3: Quantitative comparison of MDMM [Lee et al. \(2020\)](#), MDMM with MSGAN [Mao et al. \(2019\)](#) regularizer, StarGANv2 [Choi et al. \(2020\)](#), MDMM with DivAugGAN regularizer, and StarGANv2 with DivAugGAN regularizer on *alps seasonal transfer* dataset.

		FID ↓	LPIPS ↑	NDB ↓	JSD ↓
MDMM	Summer → Spring	81.40±0.34	0.0833±0.0017	20.33±2.67	0.056±0.008
	Autumn → Spring	84.63±0.60	0.0865±0.0028	20.00±1.00	0.061±0.002
	Winter → Spring	89.22±0.25	0.0869±0.0033	17.33±1.33	0.076±0.002
	Spring → Summer	73.16±0.09	0.0817±0.0015	15.67±2.67	0.039±0.005
	Autumn → Summer	64.57±0.16	0.0847±0.0022	11.00±2.00	0.033±0.005
	Winter → Summer	78.45±0.25	0.0847±0.0032	25.00±2.00	0.084±0.004
	Spring → Autumn	71.44±0.20	0.0833±0.0008	17.33±2.33	0.042±0.006
	Summer → Autumn	66.01±0.31	0.0890±0.0025	15.67±0.67	0.036±0.002
	Winter → Autumn	75.31±0.40	0.0848±0.0028	31.33±1.67	0.095±0.009
	Spring → Winter	80.64±0.14	0.0882±0.0030	24.67±2.33	0.131±0.007
MDMM +MSGAN	Summer → Winter	78.10±0.10	0.0902±0.0017	28.00±1.00	0.122±0.005
	Autumn → Winter	76.87±0.53	0.0921±0.0041	26.33±2.67	0.113±0.003
	Summer → Spring	75.38±0.19	0.1353±0.0034	14.33±0.67	0.041±0.004
	Autumn → Spring	79.54±0.23	0.1426±0.0054	17.00±3.00	0.046±0.006
	Winter → Spring	82.05±0.54	0.1352±0.0047	19.67±0.67	0.066±0.005
	Spring → Summer	69.33±0.23	0.1328±0.0021	21.00±2.00	0.051±0.005
	Autumn → Summer	60.69±0.08	0.1396±0.0016	16.00±2.00	0.038±0.001
	Winter → Summer	72.07±0.27	0.1362±0.0009	22.00±2.00	0.087±0.004
	Spring → Autumn	66.34±0.33	0.1396±0.0052	16.33±2.67	0.043±0.005
	Summer → Autumn	60.26±0.29	0.1420±0.0034	14.33±1.33	0.037±0.002
DivAug GAN(M)	Winter → Autumn	66.79±0.35	0.1386±0.0018	25.00±1.00	0.090±0.004
	Spring → Winter	73.45±0.41	0.1397±0.0016	20.33±1.33	0.089±0.004
	Summer → Winter	68.25±0.30	0.1438±0.0048	22.33±1.33	0.072±0.003
	Autumn → Winter	70.20±0.51	0.1513±0.0035	26.67±2.33	0.094±0.006
	Summer → Spring	73.81±0.22	0.1697±0.0016	11.67±1.33	0.044±0.003
	Autumn → Spring	76.03±0.23	0.1715±0.0045	18.00±1.00	0.048±0.003
	Winter → Spring	78.27±0.33	0.1803±0.0004	13.67±1.33	0.045±0.002
	Spring → Summer	65.57±0.12	0.1672±0.0036	18.00±3.00	0.047±0.004
	Autumn → Summer	55.67±0.28	0.1671±0.0007	15.67±1.33	0.040±0.003
	Winter → Summer	69.59±0.56	0.1707±0.0026	20.00±2.00	0.059±0.001
StarGANv2	Spring → Autumn	62.80±0.13	0.1706±0.0025	15.33±1.67	0.040±0.001
	Summer → Autumn	57.87±0.15	0.1684±0.0016	13.33±1.67	0.032±0.004
	Winter → Autumn	66.88±0.18	0.1729±0.0058	20.33±3.67	0.058±0.006
	Spring → Winter	68.75±0.34	0.1822±0.0024	27.00±2.00	0.112±0.003
	Summer → Winter	64.54±0.20	0.1759±0.0026	20.67±1.67	0.078±0.006
	Autumn → Winter	62.15±0.16	0.1801±0.0019	23.67±2.33	0.101±0.008
	Summer → Spring	58.51±0.38	0.2889±0.0042	16.67±1.33	0.043±0.002
	Autumn → Spring	58.12±0.54	0.3009±0.0030	17.33±1.67	0.051±0.004
	Winter → Spring	68.84±0.20	0.2966±0.0015	22.00±2.00	0.073±0.005
	Spring → Summer	46.13±0.46	0.2811±0.0058	13.33±1.33	0.028±0.002
DivAug GAN(S)	Autumn → Summer	43.30±0.24	0.2865±0.0062	10.00±3.00	0.025±0.004
	Winter → Summer	51.54±0.40	0.3019±0.0021	20.33±0.67	0.075±0.009
	Spring → Autumn	42.26±0.31	0.3005±0.0063	15.67±2.33	0.038±0.003
	Summer → Autumn	45.04±0.64	0.2934±0.0029	12.67±0.67	0.035±0.003
	Winter → Autumn	49.03±0.38	0.3223±0.0047	18.67±1.33	0.069±0.002
	Spring → Winter	48.10±0.08	0.3182±0.0018	12.33±2.67	0.038±0.004
	Summer → Winter	48.82±0.50	0.3292±0.0080	14.67±3.33	0.040±0.003
	Autumn → Winter	47.46±0.28	0.3313±0.0073	10.33±0.67	0.039±0.003
	Summer → Spring	54.57±0.36	0.4055±0.0082	15.67±1.67	0.039±0.004
	Autumn → Spring	53.44±0.68	0.4190±0.0051	12.33±1.33	0.033±0.003
DivAug GAN(S)	Winter → Spring	58.26±0.72	0.4312±0.0049	15.33±1.33	0.051±0.003
	Spring → Summer	39.02±0.24	0.4054±0.0032	13.67±2.33	0.032±0.001
	Autumn → Summer	37.56±0.31	0.4052±0.0084	10.00±2.00	0.027±0.005
	Winter → Summer	43.35±0.35	0.4277±0.0045	20.33±1.33	0.075±0.005
	Spring → Autumn	35.38±0.28	0.4295±0.0018	12.33±1.67	0.030±0.006
	Summer → Autumn	37.32±0.85	0.4216±0.0059	10.67±2.33	0.027±0.005
	Winter → Autumn	39.96±0.71	0.4625±0.0046	19.00±2.00	0.060±0.007
	Spring → Winter	42.98±0.39	0.4202±0.0059	13.00±2.00	0.035±0.002
	Summer → Winter	44.50±0.05	0.4247±0.0074	14.00±4.00	0.038±0.004
	Autumn → Winter	42.75±0.71	0.4271±0.0071	11.33±2.67	0.041±0.003



Figure 12: Qualitative comparison results on *alps seasonal transfer* dataset for multi-domain multimodal image-to-image translation. Translation results of *spring* \rightarrow *summer/autumn/winter* are illustrated. Rotate the figure 90 degrees clockwise to see.

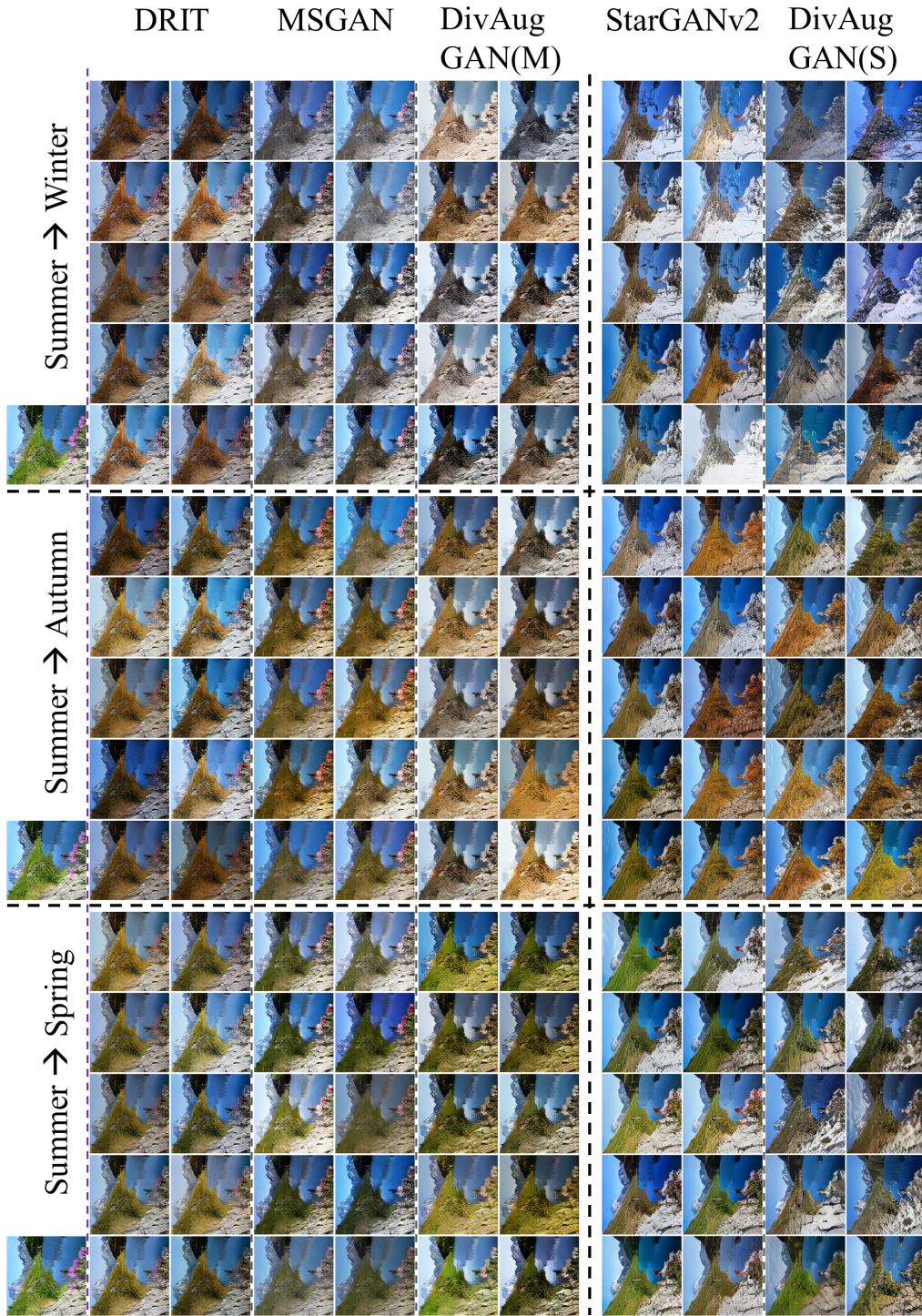


Figure 13: Qualitative comparison results on *alps seasonal transfer* dataset for multi-domain multimodal image-to-image translation. Translation results of *summer → spring/autumn/winter* are illustrated. Rotate the figure 90 degrees clockwise to see.



Figure 14: Qualitative comparison results on *alps seasonal transfer* dataset for multi-domain multimodal image-to-image translation. Translation results of *autumn* \rightarrow *spring/summer/winter* are illustrated. Rotate the figure 90 degrees clockwise to see.

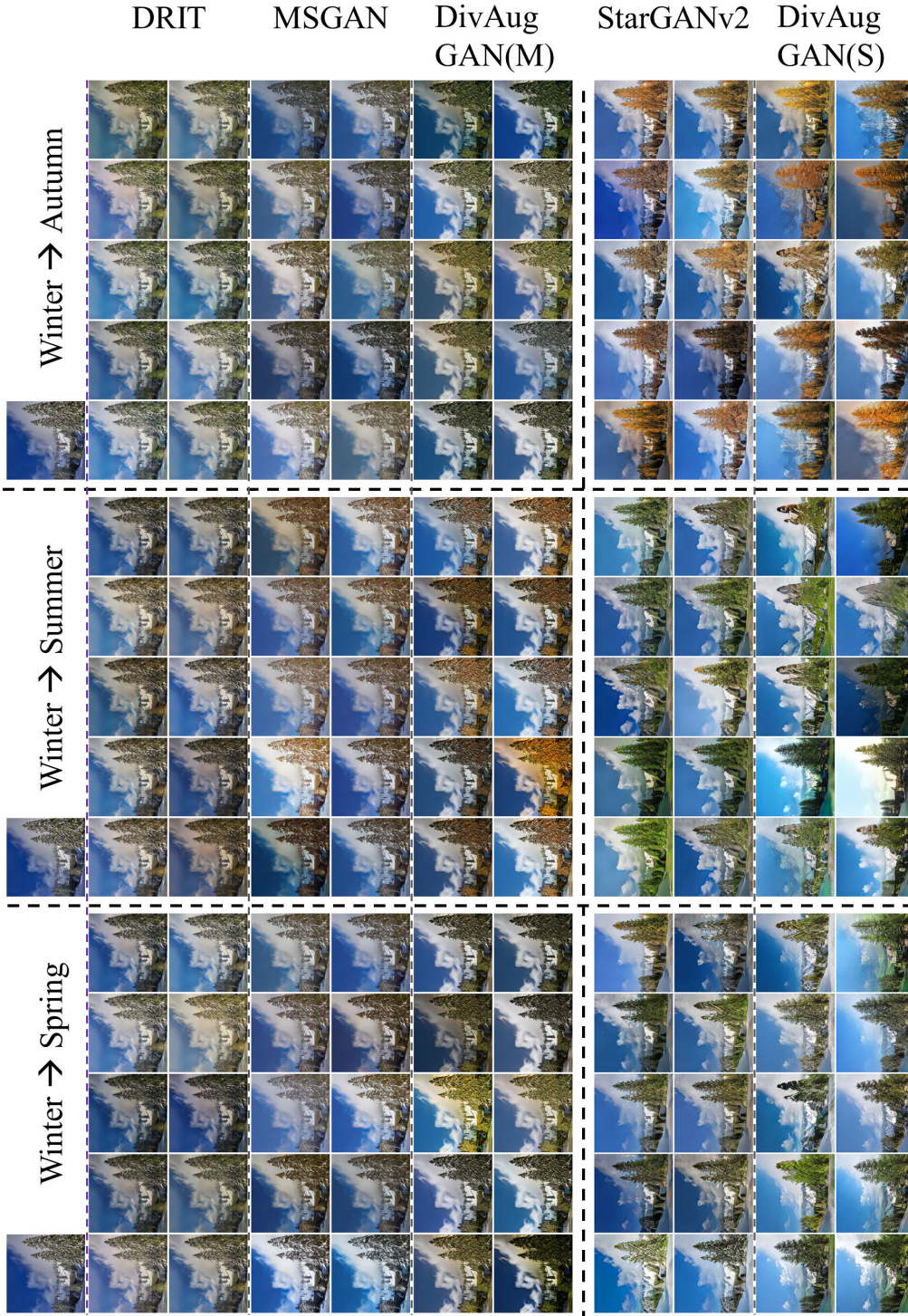


Figure 15: Qualitative comparison results on *alps seasonal transfer* dataset for multi-domain multimodal image-to-image translation. Translation results of *winter → spring/summer/autumn* are illustrated. Rotate the figure 90 degrees clockwise to see.

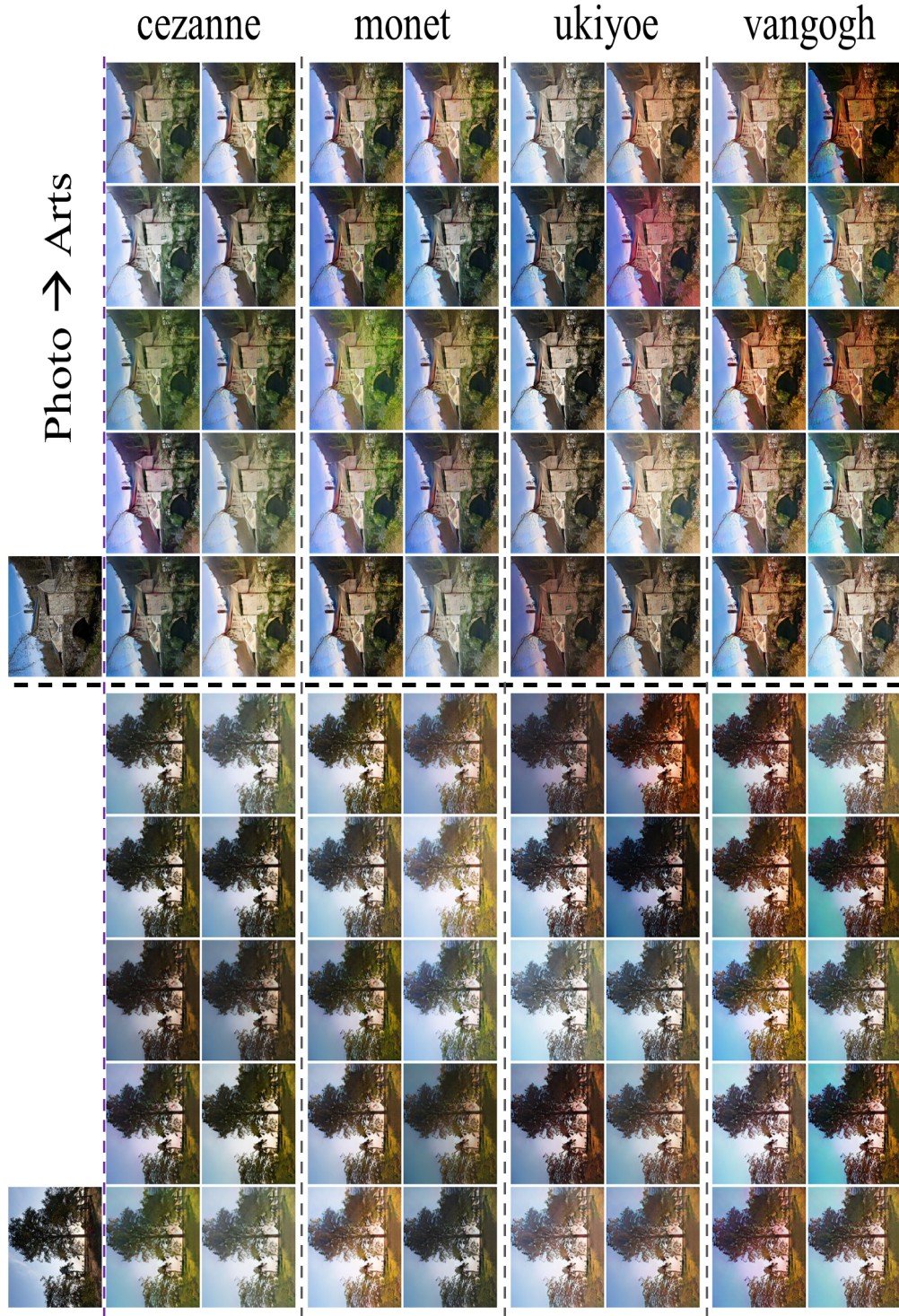


Figure 16: More qualitative results of DivAugGAN on $photo \Rightarrow \text{cezanne/monet/ukiyoe/vangogh}$ for multi-domain multimodal image-to-image translation tasks. Rotate the figure 90 degrees clockwise to see.

REFERENCES

- Asha Anoopshah, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. Combogan: Unrestrained scalability for image domain translation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018. 4
- Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. arxiv e-prints, art. In *2017 International Conference on Learning Representations (ICLR)*, 2017. 2
- Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017. 2
- Tong Che, Yanran Li, Athul Paul Jacob, Yoshua Bengio, and Wenjie Li. Mode regularized generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2016. 2
- Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 4, 5, 15, 17
- Wei-Ta Chu, Xiang-You Zheng, and Ding-Shiuan Ding. Camera as weather sensor: Estimating weather information from single images. *Journal of Visual Communication and Image Representation*, 2017. 4
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. 4
- Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning. In *International Conference on Learning Representations (ICLR)*, 2017. 2
- Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Olivier Mastropietro, Alex Lamb, Martin Arjovsky, and Aaron Courville. Adversarially learned inference. In *International Conference on Learning Representations (ICLR)*, 2017. 2
- Arnab Ghosh, Viveka Kulharia, Vinay P. Namboodiri, Philip H.S. Torr, and Puneet K. Dokania. Multi-agent diverse generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27 (NeurIPS)*. 2014. 2
- Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems 30*. 2017. 2
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *IEEE International Conference on Computer Vision (ICCV)*, December 2015. 4
- Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems 30 (NeurIPS)*. 2017. 2, 4
- Quan Hoang, Tu Dinh Nguyen, Trung Le, and Dinh Phung. Mgan: Training generative adversarial nets with multiple generators. In *International Conference on Learning Representations (ICLR)*, 2018. 2
- Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 3
- Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 3

- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NeurIPS)*. 2012. [4](#)
- Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. In *Proceedings of The 33rd International Conference on Machine Learning (ICML)*, 2016. [2](#)
- Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *The European Conference on Computer Vision (ECCV)*, 2018. [3](#), [4](#), [5](#)
- Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Singh, and Ming-Hsuan Yang. Drit++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision*, 2020. [5](#), [15](#), [17](#)
- Jianxin Lin, Zhibo Chen, Yingce Xia, Sen Liu, Tao Qin, and Jiebo Luo. Exploring explicit domain supervision for latent space disentanglement in unpaired image-to-image translation. *IEEE transactions on pattern analysis and machine intelligence*, 2019. [2](#)
- Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 29*. 2016. [2](#)
- Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. [5](#), [17](#)
- Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2017. [2](#), [15](#)
- Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *2018 International Conference on Learning Representations (ICLR)*, 2018. [2](#)
- Tu Nguyen, Trung Le, Hung Vu, and Dinh Phung. Dual discriminator generative adversarial nets. In *Advances in Neural Information Processing Systems 30*. 2017. [2](#)
- Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier GANs. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017. [2](#)
- Augustus Odena, Jacob Buckman, Catherine Olsson, Tom Brown, Christopher Olah, Colin Raffel, and Ian Goodfellow. Is generator conditioning causally related to GAN performance? In *Proceedings of the 35th International Conference on Machine Learning*, 2018. [2](#), [3](#)
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*. 2019. [3](#)
- Eitan Richardson and Yair Weiss. On gans and gmms. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*. 2018. [5](#)
- Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 2015. [4](#)
- Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems 29 (NeurIPS)*. 2016. [2](#)

- Akash Srivastava, Lazar Valkov, Chris Russell, Michael U. Gutmann, and Charles Sutton. Veegan: Reducing mode collapse in gans using implicit variational learning. In *Advances in Neural Information Processing Systems 30*. 2017. [2](#)
- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [4](#)
- Dingdong Yang, Seunghoon Hong, Yunseok Jang, Tianchen Zhao, and Honglak Lee. Diversity-sensitive conditional generative adversarial networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019. [3](#)
- Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. [4](#)
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. [4](#)