A Survey on Generative Modeling with Limited Data, Few Shots, and Zero Shot

MILAD ABDOLLAHZADEH, TOUBA MALEKZADEH*, CHRISTOPHER T.H. TEO*, KESHIGEYAN CHANDRASEGARAN*, GUIMENG LIU, and NGAI-MAN CHEUNG†,
Singapore University of Technology and Design (SUTD), Singapore

In machine learning, generative modeling aims to learn to generate new data statistically similar to the training data distribution. In this paper, we survey learning generative models under limited data, few shots and zero shot, referred to as **Generative Modeling under Data Constraint (GM-DC)**. This is an important topic when data acquisition is challenging, *e.g.* healthcare applications. We discuss background, challenges, and propose two taxonomies: one on GM-DC tasks and another on GM-DC approaches. Importantly, we study interactions between different GM-DC tasks and approaches. Furthermore, we highlight research gaps, research trends, and potential avenues for future exploration. Project website: https://gmdc-survey.github.io.

CCS Concepts: • **Computing methodologies** → **Neural networks**; **Computer vision**.

Additional Key Words and Phrases: Generative Modeling, Generative AI, Generative Adversarial Networks, Diffusion Models, Variational Auto-Encoder, Few-Shot Learning, Zero-Shot Learning, Transfer Learning, Data Augmentation.

ACM Reference Format:
Milad Abdollahzadeh, Touba Malekzadeh, Christopher T.H. Teo,, Keshigeyan Chandrasegaran, Guimeng Liu, and Ngai-Man Cheung. 2023. A Survey on Generative Modeling with Limited Data, Few Shots, and Zero Shot. 1, 1 (July 2023), 37 pages. https://doi.org/na

1 INTRODUCTION

Generative modeling is a field of machine learning that focuses on learning the underlying distribution of the training samples, enabling the generation of new samples that exhibit similar statistical properties to the training data. Generative modeling has profound impacts in various fields including computer vision [12, 78, 134], natural language processing [52, 171, 202] and data engineering [6, 76, 168]. Over the years, significant advancements have been made in generative modeling. Innovative approaches such as Generative Adversarial Networks (GANs) [7, 12, 22, 48, 77, 125, 223], Variational Autoencoders (VAEs) [83, 170, 171], and Diffusion Models (DMs) [32, 118, 140, 161] have played a pivotal role in enhancing the quality and diversity of generated samples. The advancements in generative modeling have fueled the recent disruption in generative AI, unlocking

*Equal Contribution
†Corresponding Author
[1]This research is supported by the National Research Foundation, Singapore under its AI Singapore Programmes (AISG Award No.: AISG2-RP-2021-021; AISG Award No.: AISG-100E2018-005).

Authors' address: Milad Abdollahzadeh; Touba Malekzadeh, {milad_abdollahzadeh,touba_malekzadeh}@sutd.edu.sg; Christopher T.H. Teo, christopher_teo@mymail.sutd.edu.sg;
Keshigeyan Chandrasegaran; Guimeng Liu; Ngai-Man Cheung, {keshigeyan,guimeng_liu,ngaiman_cheung}@sutd.edu.sg, Singapore University of Technology and Design (SUTD), 8 Somapah Rd, Singapore, 487372.

XXXX-XXXX/2023/7-ART $15.00
https://doi.org/na

new possibilities in various applications such as image synthesis [24, 136], text generation [56, 66], music composition [37, 190], genomics [115], and more [86, 148]. The ability to generate realistic and diverse samples has opened doors to creative applications and novel solutions [137, 142].

Research on generative modeling has been mainly focusing on setups with sizeable training datasets. StyleGAN [77] learns to generate realistic and diverse face images using Flickr-Faces-HQ (FFHQ), a high-quality dataset of 70k human face images collected from the photo-sharing website Flickr. The more recent text-to-image generative model is trained on millions of image-text pairs, e.g. Latent Diffusion Model [140] is trained on LAION-400M with 400 million samples [152]. However, in many domains (*e.g.* , medical), the collection of data samples is challenging and expensive.

In this paper, we survey Generative Modeling under Data Constraint (GM-DC). This research area is important for many domains/ applications where challenges in data collection exist. We conduct a thorough literature review on learning generative models under limited data, few shots, and zero shot. *Our survey is the first to provide a comprehensive overview and detailed analysis of all types of generative models, tasks, and approaches studied in GM-DC, offering an accessible guide on the research landscape* (Fig. 1). We cover the essential backgrounds, provide detailed analysis of unique challenges of GM-DC, discuss current trends, and present the latest advancements in GM-DC.

Our Contributions: i) Trends, technical evolution, and statistics of GM-DC (Fig. 3; Fig. 4; Sec. 5.1); ii) New insights on GM-DC challenges (Sec. 3.2); iii) Two novel and detailed taxonomies, one on GM-DC tasks (Sec. 3.1) and another on GM-DC approaches (Sec. 4); iv) A novel Sankey diagram to visualize the research landscape and relationship between GM-DC tasks, approaches, and methods (Fig. 1); v) An organized summary of individual GM-DC works (Sec. 4); vi) Discussion of future directions (Sec. 5.2). We further provide a project website with an interactive diagram to visualize GM-DC landscape. Our survey aims to be an accessible guide to provide fresh perspectives on the current research landscape, organized pointers to comprehensive literature, and insightful trends on the latest advances of GM-DC.

Survey on GM-DC is inadequate, and our work aims to fill this gap. We found only one survey in arXiv on the early work of GM-DC focusing on some aspects of GM-DC [105]. This previous survey has focused on a subset of GM-DC papers, studying only works with GANs as the generative model and a subset of technical tasks/ approaches. Our survey differentiates itself from [105] in: i) Scope - Our survey is the first to cover all types of generative models and all GM-DC tasks and approaches (Fig. 3); ii) Scale - Our study includes 113 papers and covers broad GM-DC works, while previous survey [105] covers only ≈27% of works discussed in our survey (Fig. 2); iii) Timeliness - Our survey collects and surveys the most up-to-date papers in GM-DC; iv) Detailedness - Our paper includes detailed visualizations (Sankey diagram, charts) and tables to highlight interactions and important attributes of GM-DC literature; v) Technical evolution analysis - Our paper analyzes the evolution of GM-DC tasks and approaches, providing new perspectives on recent advances; vi) Horizon analysis - Our paper discusses distinctive obstacles encountered in GM-DC and identifies avenues for future research.

The rest of the paper is organized as follows. In Sec. 2 we provide the necessary background. In Sec. 3, we discuss GM-DC tasks and unique challenges. In Sec. 4, we analyze GM-DC approaches and methods. In Sec. 5, we discuss open research problems and future directions. Sec. 6 concludes the survey.

2 BACKGROUND

In this section, we first define 'domain' and 'generative modeling', then we discuss common approaches of generative modeling and data constraints studied in GM-DC.

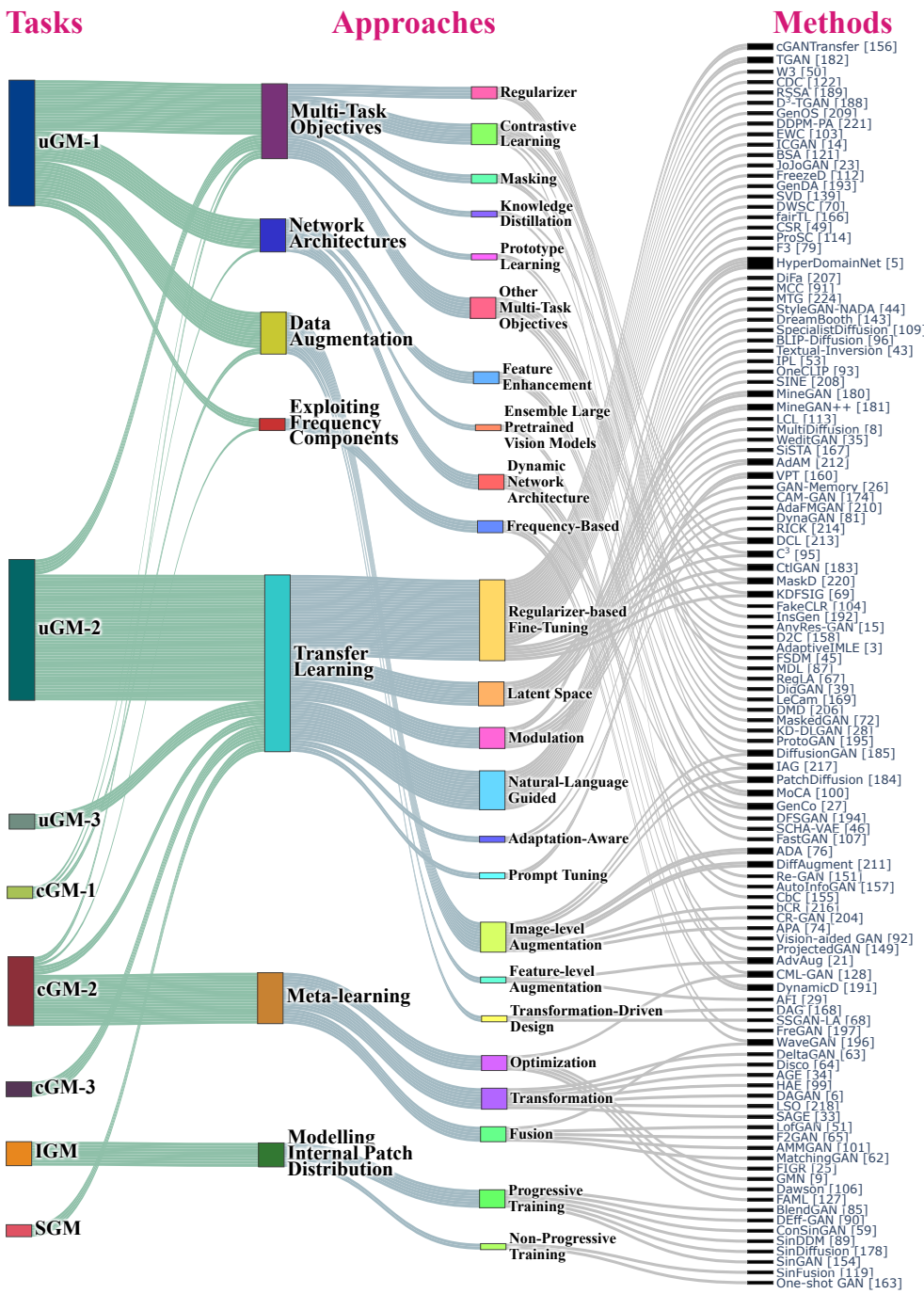


Fig. 1. **Research Landscape of GM-DC.** The figure shows the interaction between GM-DC tasks and approaches (main and sub categories), and GM-DC methods. Tasks are defined in our proposed taxonomy in Tab. 2, and approaches in our proposed taxonomy in Tab. 3. An interactive version of this diagram is available at our project website. Best viewed in color and with zoom.

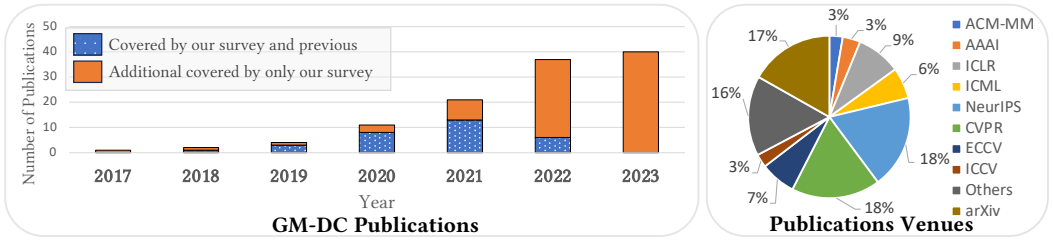


Fig. 2. Overall publications statistics in GM-DC. **GM-DC Publications (Left):** GM-DC publication trends indicate rising interest in this area. We remark that the previous survey [105] only covers ~27% of publications discussed in our survey. **Publication Venues (Right):** The distribution of publications in major machine learning and computer vision venues, other venues, and arXiv. Best viewed in color.

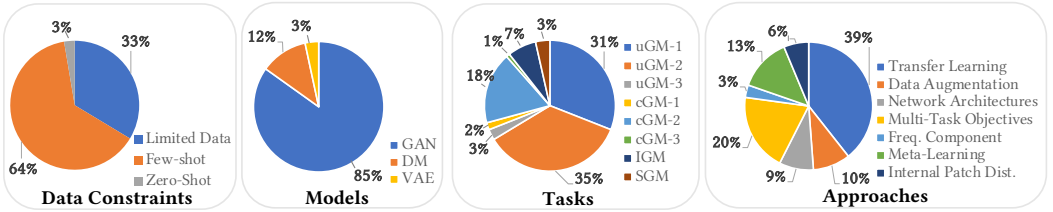


Fig. 3. Analysis of publications in GM-DC. **Data Constraints:** Different types of data constraints studied in GM-DC. See Sec. 2 for more details on setups. **Models:** Different types of models are studied including Generative Adversarial Network (GAN), Diffusion Model (DM), and Variational Auto-Encoder (VAE). **Tasks:** Different GM-DC tasks that are studied; See Sec. 3.1, and Tab. 2 for details on task definitions in our proposed task taxonomy. **Approaches:** Different approaches that are applied for addressing GM-DC; More details on our proposed taxonomy of approaches can be found in Sec. 4 and Tab. 3. Best viewed in color.

Domain. In this survey, a *domain* consists of two components: i) a sample space $\mathcal{X}$, and ii) a marginal probability distribution P_{data}, which models the probability of samples from $\mathcal{X}$ [124]. This is written as $\mathcal{D} = \{\mathcal{X}, P_{data}\}$, and $x \sim P_{data} \in \mathcal{X}$ denoting a sample in this space. An example of a domain is the domain of image of human faces: $\mathcal{D}^h = \{\mathcal{X}, P^h_{data}\}$. Here $\mathcal{X}$ is the sample space of images, and P^h_{data} is the probability distribution of human faces.

Generative Modeling. Given a set of training sample x of a domain $\mathcal{D} = \{\mathcal{X}, P_{data}\}$, i.e., with an underlying probability distribution P_{data}, generative modeling aims to learn to capture P_{data} —sometimes also denoted as $P(x)$ in literature. The result of generative modeling is a *generative model* G encoding a probability distribution P_{model}. The learning objective is to have P_{model} similar to P_{data} statistically. After the training, G can generate samples following P_{model}. For example, generative modeling with a training set of human face images aims to learn to capture P^h_{data}, thereby the resulting G^h can generate human face images that are statistically similar to samples from P^h_{data}. We also refer to the domain of training samples as *target domain*.

Conditional vs Unconditional Sample Generation. After learning the underlying distribution of data P_{data}, the generative model can generate new samples by sampling from the learned distribution P_{model}. Typically, generation starts with sampling a random vector z —also called latent code— as input. Then, this input is passed into the generative model G to transform the latent code into a new sample $G(z) \sim P_{model}$. Ideally, a good generator is able to capture the characteristics, quality and diversity of the training dataset, *i.e.*, P_{model} is similar to P_{data} statistically. If an additional

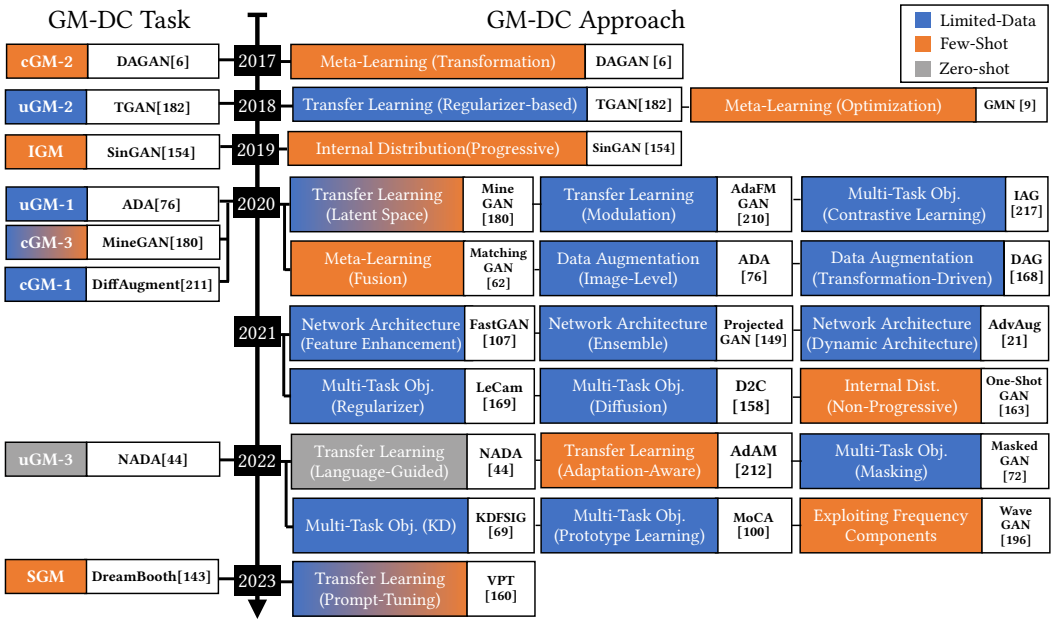


Fig. 4. Illustration of the **timeline** when a GM-DC task/approach was introduced based on our proposed taxonomies: task taxonomy (details in Sec. 3.1, and Tab. 2), and approach taxonomy (details in Sec. 4, and Tab. 3). Best viewed in color.

condition c (like a class label or attribute) is used alongside with the latent code to steer the sample generation towards c, the sample generation is called conditional generation: $G(z, c)$.

2.1 Approaches for Generative Modeling

Earlier works on generative models study Gaussian Mixture Models [138], Hidden Markov Models [129], Latent Dirichlet Allocation [20] and Boltzmann Machines [2]. With the introduction of deep neural networks, recent works study powerful generative models, particularly those for image generation, which most GM-DC works focus on.

Variational Auto Encoders (VAE) [131]. VAE is a variant of Auto-Encoder (AE) [203], where both consist of the encoder and decoder networks. AE focuses on dimensional reduction. The encoder in AE learns to map an input x into a latent (compressed) representation, $z = E(x)$. Then, the decoder aims to reconstruct the image from that latent representation, $\hat{x} = D(z)$. Model parameters are optimized with the following reconstruction loss:

$$\mathcal{L}_{rec} = ||x - D(z)||_2 \tag{1}$$

AEs are notorious for latent space irregularity making them improper for sample generation [84]. VAE aims to address this problem by enforcing E to return a normal distribution over latent space. Assuming a distribution $z \sim \mathcal{N}(\mu, \sigma^2)$ for latent space, this is done by adding the KL-divergence term to the loss function:

$$\mathcal{L} = ||x - D(z)||_2 + KL(\mathcal{N}(\mu, \sigma^2), \mathcal{N}(0, I)) \tag{2}$$

Due to the challenges of direct maximization of the likelihood in pixel space, Vector-Quantized VAE (VQ-VAE) proposes *tokenization* where a codebook $\mathbf{e}_k$, $k \in 1, \ldots, K$ is used to quantize the embeddings $E(x)$ into visual tokens (indices), acting like a lookup table. In addition, a latent prior

Table 1. List of common datasets used in GM-DC works. Number of samples (# Samples) refers to the sample size of the entire dataset. In GM-DC experiments, usually, only a subset of the dataset is used. We remark that ❍/●denotes the absence/presence of the dataset under the data constraint settings: **LD**: Limited-Data, **FS**: Few-Shot and **ZS**: Zero-Shot, and Labels indicate if training labels are available (but not necessarily used).

Dataset	Description	# Samples	Resolution	LD	FS	ZS	Labels
Flickr-Faces-HQ (FFHQ) [77]	Images with human faces, containing variation in terms of age, ethnicity, and image background.	70K	1024×1024	●	●	●	❍
Large-scale Scene Understanding (LSUN) [201]	Images with large-scale scene containing 10 scene and 20 object categories.	3M	256×256	●	●	❍	●
MetFace [76]	Images depicting paintings, drawings, and statues of human faces	1336	1024×1024	●	●	❍	❍
BreCaHAD [4]	Images of breast cancer histopathology.	162	1360×1024	●	❍	❍	❍
Animal FacesHQ (AFHQ) [22]	Images of animal faces in the domains of cat, dog, and wildlife.	15K	512×512	●	●	❍	❍
CIFAR-10 [88]	Images including objects and animals.	60K	32×32	●	❍	❍	●
CIFAR-100 [88]	A dataset similar to CIFAR-10, but with 100 classes	60K	32×32	●	❍	❍	●
100-shot Obama/ Gumpy Cat/Panda [211]	Colored images of Obama/Gumpy Cat/Panda	100	256×256	●	❍	❍	❍
Sketches [179]	Face sketches in frontal pose, normal lighting, and neutral expressions	606	256×256	❍	●	❍	❍
Sunglasses [122]	Images of human faces wearing sunglasses.	2700	256×256	❍	●	❍	❍
Babies [122]	Images of baby faces.	2500	256×256	❍	●	❍	❍
Artistic-Faces [198]	Images containing 160 artistic portraits of 16 different artists.	160	256×256	❍	●	❍	❍
Haunted houses [201]	Images of haunted houses	1K	256×256	❍	●	❍	❍
Wrecked cars [201]	Images of wrecked cars	1K	256×256	❍	●	❍	❍

of the visual tokens is predicted (usually using a transformer), and the decoder is modified to map the visual tokens into the image space.

Generative Adversarial Models (GAN) [73, 150]. GAN applies an adversarial approach to learn the distribution of data P_{data}. It consists of a generator G and a discriminator D playing a min-max game. Specifically, given the latent code z, the G learns to generate the images $G(z)$, $z \sim P_z$, where P_z is usually a Gaussian distribution. Then, D learns to distinguish the real images $x \sim P_{data}$ from the generated ones $G(z) \sim P_{model}$. The D and G are optimized by respectively maximizing and minimizing the following value function:

$$\mathcal{V}(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))] \tag{3}$$

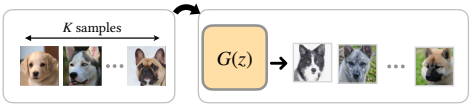
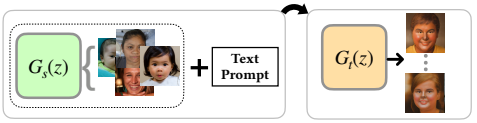
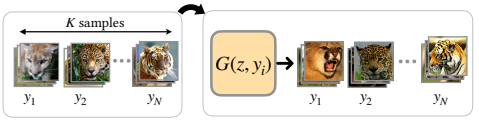
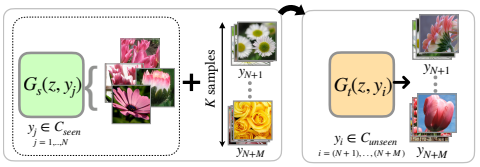
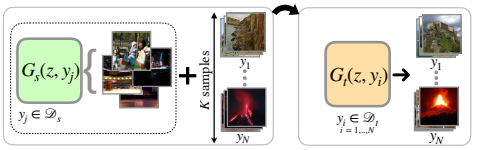
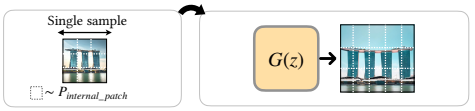
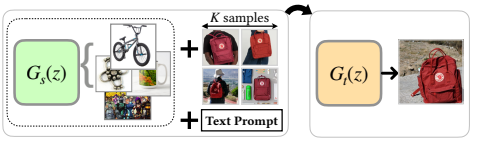
Flow-based Models [60]. The flow-based model includes a series of invertible yet differentiable functions f, between latent distribution P_z, and data distribution P_{data}. The following log-likelihood function is maximized to train $f(.|\theta)$:

$$\max_{\theta} \sum_{i=1}^{K} \log P_z(f(x^{(i)}|\theta)) + \log|\det Df(x^{(i)}|\theta)| \tag{4}$$

For ease of discussion, we simplify the model as a single flow and denote the training samples with $\{x^{(i)}\}_{i=1}^{K}$, and the Jacobian of $f(x)$ as $Df(x)$. We remark that, unlike VAEs that estimate the lower bounds of the log-likelihood, flow-based models evaluate the exact log-likelihood in their loss function.

Diffusion Models (DM) [82]. DM leverages the concept of the diffusion process from stochastic calculus and consists of forward diffusion and reverse diffusion processes. In the forward diffusion process, based on the foundations of Markov chains, the noise $\epsilon \sim \mathcal{N}(0, I)$ is iteratively added to data samples until it approaches an isotropic Gaussian distribution. Then, in the backward process, the DM learns to denoise the noisy vector x_T and reconstruct the data samples x_0. This is done by

Table 2. **Our proposed taxonomy for tasks in GM-DC**. For each task, we extract their key characteristics. [Attributes] **C**: Conditional generation, **P**: Pre-trained generator given, **I**: Images (as input), **TP**: Text-Prompt (as input), **X**: X(Cross)-domain adaptation; [Data Constraint] **LD**: Limited-Data, **FS**: Few-Shot, **ZS**: Zero-Shot. ❍/●denotes the absence/presence, respectively. Best viewed in color.

Task	Attributes					Data Constraint			Task Illustration
	C	P	I	TP	X	LD	FS	ZS	
uGM-1	❍	❍	●	❍	❍	●	❍	❍	
	Description: Given K samples from a domain $\mathcal{D}$, learn to generate diverse and high-quality samples from $\mathcal{D}$ **Example:** ADA [76] learns a StyleGAN2 using 1k images from AFHQ-Dog								
uGM-2	❍	●	●	❍	●	●	●	❍	
	Description: Given a pre-trained generator on a source domain $\mathcal{D}_s$ and K samples from a target domain $\mathcal{D}_t$, learn to generate diverse and high-quality samples from $\mathcal{D}_t$ **Example:** CDC [122] adapts a pre-trained GAN on FFHQ to Sketches using 10 samples								
uGM-3	❍	●	❍	●	●	❍	❍	●	
	Description: Given a pre-trained generator on a source domain $\mathcal{D}_s$ and a text prompt describing a target domain $\mathcal{D}_t$, learn to generate diverse and high-quality samples from $\mathcal{D}_t$ **Example:** NADA [44] adapts pre-trained GAN on FFHQ to the painting domain using *'Fernando Botero Painting'* as input								
cGM-1	●	❍	●	❍	❍	●	❍	❍	
	Description: Given K samples with class labels from a domain $\mathcal{D}$, learn to generate diverse and high-quality samples conditioning on the class labels from $\mathcal{D}$ **Example:** CbC [155] trains conditional generator on 20 classes of ImageNet Carnivores using 100 images per class								
cGM-2	●	●	●	❍	❍	❍	●	❍	
	Description: Given a pre-trained generator on the seen classes C_{seen} of a domain $\mathcal{D}$ and K samples with class labels from unseen classes C_{unseen} of $\mathcal{D}$, learn to generate diverse and high-quality samples conditioning on the class labels for C_{unseen} from $\mathcal{D}$ **Example:** LoftGAN [51] learns from 85 classes of Flowers to generate images for an unseen class with only 3 samples								
cGM-3	●	●	●	❍	●	●	●	❍	
	Description: Given a pre-trained generator on a source domain $\mathcal{D}_s$ and K samples with class labels from a target domain $\mathcal{D}_t$, learn to generate diverse and high-quality samples conditioning on the class labels from $\mathcal{D}_t$ **Example:** VPT [160] adapts a pre-trained conditional generator on ImageNet to Places365 with 500 images per class								
IGM	❍	❍	●	❍	❍	❍	●	❍	
	Description: Given K samples (usually $K = 1$) and assuming rich internal distribution for patches within these samples, learn to generate diverse and high-quality samples with the same internal patch distribution **Example:** SinDDM [89] trains a generator using a single image of Marina Bay Sands, and generates variants of it								
SGM	❍	●	●	●	❍	❍	●	❍	
	Description: Given a pre-trained generator, K samples of a particular subject, and a text prompt, learn to generate diverse and high-quality samples containing the same subject **Example:** DreamBooth [143] trains a generator using 4 images of a particular backpack and adapts it with a text-prompt to be in the *'grand canyon'*								

learning the noise estimation model ϵ_θ with minimizing the following loss function [61]:

$$\mathcal{L} = \mathbb{E}_{t,x_0,\epsilon}[||\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)||_2] \tag{5}$$

Then, during the generation process, DM first samples a noise $x_T \sim \mathcal{N}(0, I)$, and utilizes the learned noise function ϵ_θ to iteratively apply the following denoising process [61]:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}}}\epsilon_\theta(x_t, t)) + \sqrt{\beta_t}\epsilon, \quad t \in [0, T] \tag{6}$$

Here, x_t is the generated sample at step $T-t$, β_t is variance scheduler, $\alpha_t = 1-\beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

Remark. We remark that among discussed models, only GANs, DMs, and VAEs are adopted in the context of GM-DC.

2.2 Data Constraints and Commonly Used Datasets

In GM-DC, three data constraints have been considered in most works: (i) *Limited data (LD)*, when 50 to 5,000 training samples are given; (ii) *Few-Shot (FS)*, when 1 to 50 training samples are given; (iii) *Zero-Shot (ZS)*, when no training samples are given. Training under these data constraints often results in various problems *e.g.* over-fitting. We remark that these ranges are the typically used values as there are no fixed definitions in the literature. Tab. 1 lists the most common datasets used in GM-DC with related details.

3 GENERATIVE MODELING UNDER DATA CONSTRAINT: TASK TAXONOMY, CHALLENGES

In this section, first, we present our proposed taxonomy on different GM-DC tasks (Sec. 3.1) highlighting their relationships and differences based on their attributes, e.g. unconditional or conditional generation. Then, we present the unique challenges of GM-DC (Sec. 3.2), including new insights such as domain proximity, and incompatible knowledge transfer. Later, in Sec. 4, we present our proposed taxonomy on approaches for GM-DC, with a detailed review of individual work organized under our proposed taxonomy.

3.1 Generative Modeling under Data Constraint: A Taxonomy on Tasks

The goal of GM-DC is to learn to generate diverse and high-quality samples given only a small number of training samples. A number of GM-DC setups have been studied in different works (Fig. 4). In this section, we propose a **GM-DC task taxonomy** to categorize setups in different works. Tab. 2 tabulates our GM-DC task taxonomy.

(1) **Unconditional generative modeling under data constraint (uGM-1).**

Definition 1 (uGM-1)**.** *Given K samples from domain $\mathcal{D}$, learn to generate diverse and high-quality samples from $\mathcal{D}$.*

Without leveraging other side information, existing work has studied uGM-1 under limited samples ranging from 100 to several thousands. uGM-1 is an important task, especially for a domain that is distant from common domains, e.g. medical images which are distant from common personal photos in terms of content and characteristics. In such scenarios, leveraging from common domains would not provide any advantage.

(2) **Unconditional generative modeling under data constraint with pre-trained generator and cross-domain adaptation** (uGM-2).

Definition 2 (uGM-2)**.** *Given a pre-trained generator on a source domain $\mathcal{D}_s$ (with numerous and diverse samples) and K samples from a target domain $\mathcal{D}_t$, learn to generate diverse and high-quality samples from $\mathcal{D}_t$.*

uGM-2 is similar to uGM-1, except that a pre-trained generator on another source domain $\mathcal{D}_s$ is additionally given. uGM-2 is a major task in GM-DC and has been studied in many works. In most works, close proximity in semantic between $\mathcal{D}_s$ and $\mathcal{D}_t$ is assumed, *e.g.* $\mathcal{D}_s$ is photos of human faces, $\mathcal{D}_t$ is sketches of human faces. For uGM-2, transfer learning has been a popular approach to tackle this task driving GM-DC into the few-shot regime, *e.g.* only 10 samples from $\mathcal{D}_t$ are given [103] (See Sec. 4 for the taxonomy of GM-DC approaches). Recent work has started to look into the challenging setup when $\mathcal{D}_s$ and $\mathcal{D}_t$ are more semantically apart [212], *e.g.* $\mathcal{D}_s$ is photos of human faces, $\mathcal{D}_t$ is photos of cat faces. See Sec. 3.2 for further discussion on domain proximity in GM-DC.

(3) **Unconditional generative modeling under data constraint with pre-trained generator and cross-domain adaptation, using text prompt (uGM-3).**

Definition 3 (uGM-3)**.** *Given a pre-trained generator on a source domain $\mathcal{D}_s$ (with numerous and diverse samples) and a text prompt describing a target domain $\mathcal{D}_t$, learn to generate diverse and high-quality samples from $\mathcal{D}_t$.*

uGM-3 is similar to uGM-2, except that a text prompt is provided to describe $\mathcal{D}_t$ instead of samples from $\mathcal{D}_t$. Particularly, this task requires generating samples from $\mathcal{D}_t$ without seeing any sample from that domain, *i.e.* zero-shot domain adaptation. Important work to tackle this task leverages recent large vision-language models to provide textual direction to guide the adaptation of the pre-trained generator to $\mathcal{D}_t$ [44].

(4) **Conditional generative modeling under data constraint (cGM-1).**

Definition 4 (cGM-1)**.** *Given K samples with class labels from a domain $\mathcal{D}$, learn to generate diverse and high-quality samples conditioning on the class labels from $\mathcal{D}$.*

cGM-1 is similar to uGM-1 but focuses on conditional generation, *i.e.* inputs to the generator include a random latent vector and a class label. Conditional generative models such as BigGAN [12] could achieve high-quality image generation when they are trained on large-scale datasets *e.g.* ImageNet. However, under limited data, it is challenging to achieve diverse and high-quality conditional sample generation. As a natural extension of uGM-1, data augmentation has been studied for cGM-1 among other approaches, see Sec. 4.

(5) **Conditional generative modeling under data constraint with pre-trained generator (cGM-2).**

Definition 5 (cGM-2)**.** *Given a pre-trained generator on the seen classes C_{seen} of a domain $\mathcal{D}$, and K samples with class labels from unseen classes C_{unseen} of $\mathcal{D}$, learn to generate diverse and high-quality samples conditioning on the class labels for C_{unseen} from $\mathcal{D}$.*

cGM-2 is similar to cGM-1, except that a pre-trained generator on the seen classes C_{seen} is additionally given. Note that in cGM-2, C_{seen} and C_{unseen} contain disjoint classes, but both of them are from the same domain $\mathcal{D}$. For example, [156] studies the setup when CIFAR100 [88] is partitioned into 80 seen classes for the pre-trained generator and 20 unseen classes as the target, with 100 samples per unseen class given for training. Meta-learning and transfer learning (regularizer-based fine-tuning, etc.) have been effective approaches for cGM-2, see Sec. 4.

(6) **Conditional Generative Modeling under data constraint with pre-trained generator and cross-domain adaptation (cGM-3).**

Definition 6 (cGM-3)**.** *Given a pre-trained generator on a source domain $\mathcal{D}_s$ (with numerous and diverse samples) and K samples with class labels from a target domain $\mathcal{D}_t$, learn to generate diverse and high-quality samples conditioning on the class labels from $\mathcal{D}_t$.*

Table 3. **Our proposed taxonomy for approaches in GM-DC.** For each approach, the addressed GM-DC tasks (see Tab. 2 for task definitions) and the data constraints are indicated. A detailed list of methods under each sub-category is also tabulated (some methods are under multiple categories). ❍/● denotes the absence/presence of the tasks commonly addressed by each approach, and the data constraints usually considered: **LD**: Limited-Data, **FS**: Few-Shot and **ZS**: Zero-Shot.

Transfer Learning (Sec. 4.1)
Description: Improve GM-DC on target domain by knowledge of a generator pre-trained on source domain (with numerous and diverse samples).
Task: uGM-1 ❍ uGM-2 ● uGM-3 ● cGM-1 ❍ cGM-2 ● cGM-3 ●IGM ❍ SGM ● **Data constraint:** LD ● FS ● ZS ●
1) Regularizer-based Fine-Tuning: Explore regularizers to preserve source generators' knowledge.
Methods: TGAN[182], BSA[121], FreezeD[112], EWC[103], CDC[122], cGANTransfer[156], W^3[50], C^3[95], DCL[213], RSSA[189], fairTL[166], GenOS[209] , SVD[139], D^3-TGAN[188], JoJoGAN[23], KDFSIG[69], CtlGAN[183], ICGAN[14], MaskD[220], F^3[79], ICGAN [14], DDPM-PA [221], DWSC [70], CSR [49], ProSC [114]
2) Latent Space: Explore latent space of source generator to identify suitable knowledge for adaptation.
Methods: MineGAN[180], MineGAN++[181], LCL[113], WeditGAN[35], GenDA[193], SiSTA [167], MultiDiffusion [8]
3) Modulation: Leverage trainable modulation weights on top of frozen weights of the source generator.
Methods: AdaFMGAN[210], GAN-Memory [26], CAM-GAN[174], AdAM[212], DynaGAN[81], HyperDomainNet[5]
4) Natural Language-guided: Use the feedback of vision-language models to adapt the source generator with text prompts.
Methods: StyleGAN-NADA[44], MTG[224], HyperDomainNet[5], DiFa[207], OneCLIP[93], IPL[53], SINE[208], DreamBooth[143], MCC[91], Textual-Inversion[43], SpecialistDiffusion[109], BLIP-Diffusion[96]
5) Adaptation-Aware: Preserve the source generator's knowledge that is important to the adaptation task.
Methods: AdAM[212], RICK[214]
6) Prompt Tuning: Freeze the source generator and add/ generate visual prompts to guide generation for the target domain.
Methods: VPT [160]
Data Augmentation (Sec. 4.2)
Description: Improve GM-DC by increasing coverage of the data distribution by applying various transformations on the given samples.
Task: uGM-1 ● uGM-2 ❍ uGM-3 ❍ cGM-1 ❍ cGM-2 ❍ cGM-3 ❍ IGM ❍ SGM ❍ **Data constraint:** LD ● FS ❍ ZS ❍
1) Image-Level Augmentation: Apply data transformations on image space.
Methods: ADA [76], DiffAugment[211], IAG[217], DiffusionGAN[185], bCR[216], CR-GAN[204], APA [74], PatchDiffusion[184]
2) Feature-Level Augmentation: Apply data transformations on the feature space.
Methods: AdvAug[21], AFI[29]
3) Transformation-Driven Design: Leverage the information of individual transformations to design an efficient learning mechanism.
Methods: DAG[168], SSGAN-LA[68]
Network Architectures (Sec. 4.3)
Description: Design specific architecture for the generator to improve its learning under data constraints.
Task: uGM-1 ● uGM-2 ❍ uGM-3 ❍ cGM-1 ● cGM-2 ❍ cGM-3 ❍ IGM ❍ SGM ❍ **Data constraint:** LD ● FS ❍ ZS ❍
1) Feature Enhancement: Design additional modules/ layers to enhance/ retain the feature maps of the generator for better generative modeling.
Methods: FastGAN[107], MoCA[100], DFSGAN[194], SCHA-VAE [46]
2) Ensemble Large Pre-trained Vision Models: Improve architecture by integrating pre-trained vision models to enable more accurate GM-DC.
Methods: Vision-aided GAN[92], ProjectedGAN [149]
3) Dynamic Network Architecture: Improve generative learning with limited data by evolving the generator architecture during training.
Methods: CbC[155], DynamicD[191], AdvAug[21], Re-GAN[151], AutoInfoGAN [157]
Multi-Task Objectives (Sec. 4.4)
Description: Introduce additional task(s) to extract generalizable representations that are useful for all tasks, to reduce overfitting under data constraints.
Task: uGM-1 ● uGM-2 ● uGM-3 ❍ cGM-1 ● cGM-2 ❍ cGM-3 ❍ IGM ❍ SGM ❍ **Data constraint:** LD ● FS ● ZS ❍
1) Regularizer: Add an additional task objective as a regularizer to prevent an undesirable behaviour during training generative model.
Methods: LeCam[169], DigGAN[39], MDL[87], RegLA[67]
2) Contrastive Learning: Introduce a pretext task to enhance the learning process of the generative model.
Methods: InsGen[192], FakeCLR[104], DCL[213], C^3[95], ctlGAN[183], IAG[217], CML-GAN[128]
3) Masking: Mask a part of the image/ information to increase the task hardness and prevent learning the trivial solutions.
Methods: MaskedGAN[72], MaskD[220], DMD [206]

4) Knowledge Distillation: Add a task objective that enforces the generator to follow a strong teacher.
Methods: KD-DLGAN[28], KDFSIG[69]

5) Prototype Learning: Emphasize learning prototypes for samples/ concepts within the distribution as an additional task objective.
Methods: ProtoGAN[195], MoCA[100]

6) Other Multi-Task Objectives: Apply other types of multi-task objectives including co-training, patch-level learning, and diffusion.
Methods: GenCo[27], PatchDiffusion[184], AnyRes-GAN[15] , DiffusionGAN[185], D2C[158], AdaptiveIMLE [3], FSDM [45]

Exploiting Frequency Components (Sec. 4.5)

Description: Exploit frequency components to improve learning the generative model by reducing frequency bias.

Task: uGM-1 ● uGM-2 ❍ uGM-3 ❍ cGM-1 ❍ cGM-2 ● cGM-3 ❍ IGM ❍ SGM ❍ **Data constraint:** LD ● FS ● ZS ❍

Methods: FreGAN[197], WaveGAN[196], MaskedGAN[72], Gen-co[27]

Meta-Learning (Sec. 4.6)

Description: Learn meta-knowledge from seen classes to improve generator learning for unseen classes.

Task: uGM-1 ❍ uGM-2 ❍ uGM-3 ❍ cGM-1 ❍ cGM-2 ● cGM-3 ❍ IGM ❍ SGM ❍ **Data constraint:** LD ❍ FS ● ZS ❍

1) Optimization: Learn initialization weights from the seen classes as meta-knowledge to enable quick adaptation to unseen classes.
Methods: GMN[9], FIGR[25], Dawson[106], FAML[127], CML-GAN[128]

2) Transformation: Learn sample transformations from the seen classes as meta-knowledge and use them for sample generation for unseen classes.
Methods: DAGAN[6], DeltaGAN[63], Disco[64], AGE[34], SAGE[33], HAE[99], LSO [218]

3) Fusion: Learn to fuse the samples of the seen classes as meta-knowledge, and apply learned meta-knowledge to generation for unseen classes.
Methods: MatchingGAN[62], F2GAN[65], LofGAN[51], WaveGAN[196], AMMGAN[101]

Modeling Internal Patch Distribution (Sec. 4.7)

Description: Learn the internal patch distribution within one image to generate diverse samples with the same visual content (patch distribution).

Task: uGM-1 ❍ uGM-2 ❍ uGM-3 ❍ cGM-1 ❍ cGM-2 ❍ cGM-3 ❍ IGM ● SGM ❍ **Data constraint:** LD ❍ FS ● ZS ❍

1) Progressive Training: Train a generative model progressively to learn the patch distribution at different scales/ noise levels.
Methods: SinDiffusion[178], SinDDM[89], Deff-GAN[90], BlendGAN[85], SinGAN[154], ConSinGAN[59]

2) Non-progressive Training: Train a generative model on the same scale/ noise but with changes to the model's architecture.
Methods: SinFusion[119], One-Shot GAN[163]

cGM-3 is similar to uGM-2 as cross-domain adaptation is required in both tasks, but cGM-3 focuses on conditional generation while uGM-2 focuses on unconditional generation. Furthermore, cGM-3 is similar to cGM-2, but seen classes and unseen classes are from different domains in cGM-3. For example, [156] has studied the setup when a pre-trained generator on ImageNet is adapted to generate samples for several classes from Places365 [219]. Transfer learning is one of the effective approaches for cGM-3, see Sec. 4.

(7) **Internal patch distribution Generative Modeling (IGM).**

Definition 7 (IGM)**.** *Given K samples and assuming rich internal distribution for patches within these samples, learn to generate diverse and high-quality samples with the same internal patch distribution.*

IGM aims to capture the internal distribution of patches within the samples. With the model capturing the samples' patch statistics, it is then possible to generate high quality, diverse samples with the same content as the given training samples. In most works, $K = 1$, and IGM focuses on images [154], learning to generate new images with significant variability while maintaining both the global structure and fine textures of the training image.

(8) **Subject-driven Generative Modeling (SGM).**

Definition 8 (SGM)**.** *Given K samples of a particular subject and a text prompt, learn to generate diverse and high-quality samples containing the same subject.*

SGM is a recent GM-DC task introduced in [143]. Given a few images (3-5 in most cases) of a subject and leveraging a large text-to-image generative model, [143] learns to generate diverse images of the subject in different contexts with the guidance of text prompts. The goals are: i) to

achieve natural interactions between the subject and diverse new contexts, and ii) to maintain high fidelity to the key visual features of the subject. In [143], a natural language-guided transfer learning approach and a new prior preservation loss have been proposed to achieve SGM.

3.2 Generative Modeling under Data Constraint: Challenges

3.2.1 Challenges for Training Generative Models under Data Constraint. Data constraints typically introduce additional challenges and amplify existing ones when training generative models. Here, we delve into the challenges of training GM-DC. These limitations include pervasive issues of overfitting and frequency bias which are commonly observed across various approaches. Additionally, knowledge transfer between domains brings forth specific problems including the proximity between source and target domains and the transfer of incompatible source knowledge. As shown in Fig. 3, around 39% of works directly rely on knowledge transfer as a mainstream method to tackle GM-DC, and more than 20% of works propose methods based on other approaches that are compatible with transfer learning.

Overfitting to Training Data. In machine learning, overfitting is a common issue when powerful models start to memorize the training data instead of learning the generalizable semantics [147]. In generative modeling, the overfitting problem exacerbates under data constraints due to the high capacity of current generative models [76, 107, 121]. When limited training data is available, generative models may simply remember the training data [103, 122] and learn to generate the exact training samples [212] instead of capturing the data distribution. Furthermore, under data constraints, generative modeling is more prone to mode collapse [168], i.e., the generators learn only a limited set of modes and fail to capture other modes of the data distribution, resulting in limited diversity in generated samples [116, 200].

Frequency Biases. Generative models are notorious for their spectral bias [80, 133], *i.e.* tendency to prioritize fitting low-frequency components while disregarding high-frequency components within a data distribution [17, 36, 165]. The exclusion of these high-frequency components which encode intricate image details [47] can significantly impact the quality of generated samples, *i.e.* , accurate modeling of high-frequency details is critical in various fields including medical imaging (X-rays, CT-scans, MRIs), satellite/ aerial imaging, astrophotography, and art restoration. This issue becomes more severe under limited data [196, 197] as even advanced network structures tailored for such scenarios [107] struggle to maintain the desired level of details in generated samples.

Modeling Distant/ Remote Target Domains under GM-DC Setups. Substantial number of GM-DC tasks rely on the transfer learning principle (uGM-2, uGM-3, cGM-2, cGM-3, SGM), which aims to enhance the generative capabilities for a target domain by leveraging the knowledge of a generator pre-trained on a large and diverse source domain (See Fig. 1). A significant amount of research has been focused on target domains that are semantically/ perpetually similar to the source domain, *e.g.* , learn to generate Baby faces using a pre-trained generator trained on Human faces. In particular, when dealing with GM-DC setups involving significant domain shifts between the source and target domains (Human Faces→Animal Faces), many proposed methods fail to outperform a basic fine-tuning approach [212]. This is due to these methods prioritizing knowledge preservation from the source domain/ task, overlooking the adaptation step to the target domain [212]. Recently, adaptation-aware algorithms have characterized source→target domain proximity [212] and addressed GM-DC setups with pronounced domain shifts between the source and target domains (Human Faces→Animal Faces) [212, 214]. To understand the concept of distant/ remote target domains, we additionally introduce two remote target domains that further exhibit a considerable degree of domain shifts: i) Human Faces (FFHQ) [77]→ Flowers [120], ii) Human

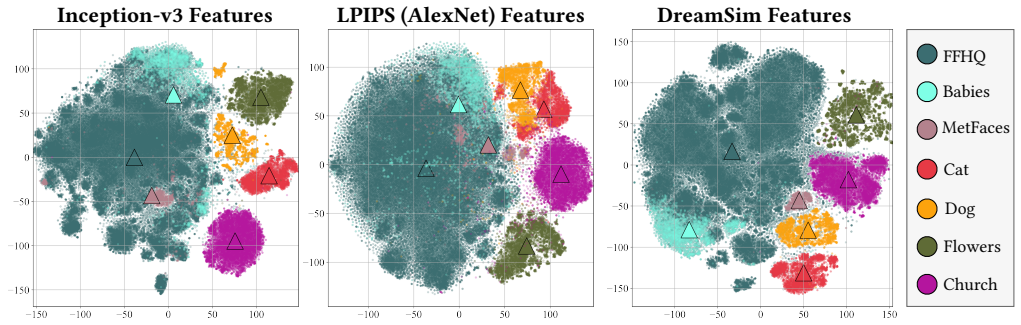


Fig. 5. **Source-target domain proximity visualization indicates that distant/ remote target domains have not been explored in GM-DC setups and are very challenging.** We use FFHQ [77] as the source domain. We show source-target domain proximity qualitatively by visualizing **Inception-v3 (Left)** [164], **LPIPS (Middle)** [205] and **DreamSim (Right)** [42] features. For feature visualization, we use t-SNE [172] and show centroids (△) for all domains. We clearly show using feature visualizations that additional setups – Flowers [120] and Church [201] – represent target domains that are remote from the source domain (FFHQ) compared to target domains used in the literature. This indicates that the exploration of distant/ remote target domains under GM-DC setups has not been pursued and poses notable challenges (Fig. 6). Best viewed in color.

Faces (FFHQ) [77]→Church [201]. Domain proximity visualization is shown in Fig. 5. In particular, we conducted a GM-DC experiment (uGM-2) to adapt a pre-trained Human face (FFHQ) generator to Flowers under 10-shot setup using AdAM [212], obtaining a FID value of 124.46. Adaptation results are shown in Fig. 6. As one can observe, multiple instances of low quality synthesis are observed in AdAM [212]. In summary, we remark that modeling distant/ remote target domains remains an important and challenging area for GM-DC.

Identifying and Removing Incompatible Knowledge Transfer. Another challenge with leveraging source domain's knowledge for GM-DC tasks is incompatible knowledge transfer, which is discovered recently [214]. In particular, many methods may transfer knowledge that is incompatible with the target domain, *e.g.* hat from source domain FFHQ to target domain flowers, significantly degrading the realisticness of the generated samples. In Fig. 6, we show multiple examples of incompatible knowledge transfer using AdAM for 10-shot flower adaptation. Although some recent effort has been invested in identifying and proactively truncating incompatible knowledge transfer [214] in Human Faces → Animal Faces adaptation setups, it is worth noting that identifying and removing incompatible knowledge remains a critical and demanding area in GM-DC.

3.2.2 Challenges on Selecting Samples for GM-DC. Although considerable research effort has been invested in developing algorithms for GM-DC, the task of sample selection for GM-DC remains a challenging and relatively unexplored area. It is essential that the samples selected for GM-DC should represent the target domain. In particular, we observe significant variation in performance with different selection of target samples as the training datasets in GM-DC. We perform a 10-shot *data-centric* GM-DC experiment using AdAM [212] to emphasize the importance of sample selection in GM-DC. Following [212, 214], we use AFHQ-Cat dataset [22] and select 3 random sets of 10-shot cat data for GM-DC. Data and 10-shot adaptation FID results are shown in Fig. 7. We obtain FID values of 90.0, 71.6 and 49.9 for Sets 1, 2 and 3 respectively (iteration=2500). This study provides evidence that sample selection plays a vital role in determining the capabilities of GM-DC. Specifically, due to cost/ privacy concerns, the role of sample selection is critical in applications including biomedical imaging, satellite/ aerial imaging and remote sensing. In summary, sample

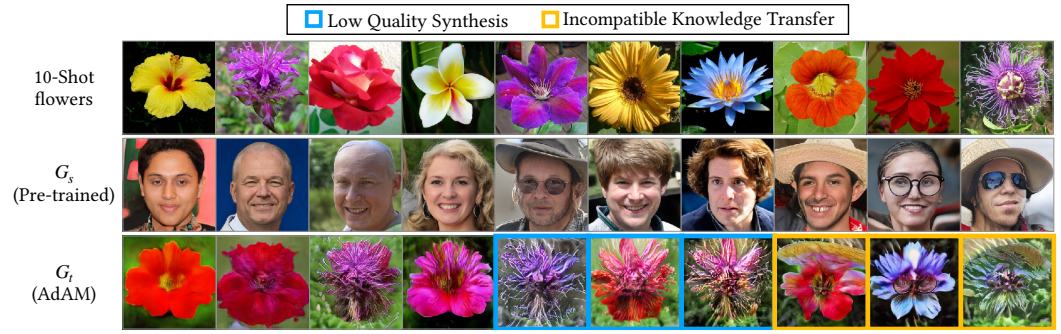


Fig. 6. **Knowledge transfer under distant/ remote target domain (Human face → Flowers) suffers from low synthesis quality and incompatible knowledge transfer.** We show 10-shot adaptation results for FFHQ → Flowers using AdAM [212]. The FID for the 10-shot adaptation using AdAM is 124.46. We highlight multiple instances of low synthesis quality and incompatible knowledge transfer (*i.e.* glasses, hat from FFHQ to flowers), showing that GM-DC modeling of remote target domains poses significant challenges. Best viewed in color.

Fig. 7. **Sample selection for GM-DC remains challenging and relatively unexplored.** We use AdAM [212] to adapt a pre-trained StyleGAN2 on FFHQ to AFHQ-Cat dataset [22] using 3 random sets of 10-shot data *(Right)*. We report FID results during training *(Left)* for these sets of data. Following [212] FID is measured between 5000 generated samples and the entire AFHQ-Cat dataset consisting of 5115 samples. We use clean-FID library [126], obtaining FID values of 90.0, 71.6 and 49.9 for Sets 1, 2, and 3 respectively at iteration=2500. As indicated by FID trends, the generative capabilities of GM-DCs are drastically influenced by sample selection. Best viewed in color.

selection for GM-DC holds significant importance and remains an area with limited investigation thus far.

3.2.3 Challenges in Evaluating Generative Models under Data Constraint. The assessment of generative modeling capabilities presents lots of challenges, encompassing both objective and subjective evaluation [94]. These issues are aggravated under low-data regimes resulting in the evaluation of GM-DC to be challenging and an active topic of research. In contemporary GM-DC literature, sample quality and diversity are used as the main attributes for evaluating generation capability. A summary of prominent metrics for GM-DC is included in Tab. 4.

Existing GM-DC evaluation metrics present multiple challenges: i) Statistical measures including FID, KID, IS, FID_{CLIP} lose their significance when dealing with setups where there is an extreme scarcity (Few-shots) or complete absence (Zero-shot) of target domain data. For example, when the reference distribution contains only 10 real images, the mean and trace components of FID are not statistically significant. ii) Although human judgment/ user feedback is used for the subjective

Table 4. List of common metrics used for evaluating GM-DC works. **LD**: Limited-Data, **FS**: Few-Shot, **ZS**: Zero-Shot. ❍/●denotes the absence/presence, respectively.

Metrics	FID [57]/ FID_{CLIP} [94]	KID [10]	IS [145]	Intra-LPIPS [122]	SIFID [154]	Image/ Text Similarity [43]	User Feedback
LD	●	●	●	❍	❍	❍	●
FS	●	●	●	●	●	●	●
ZS	❍	❍	❍	●	●	●	●

evaluation of GM-DC, the absence of a unified framework/ protocol for such evaluation strategy results in inadequacy when comparing the generative capabilities of different GM-DC models. iii) The over-reliance on objective GM-DC measures on deep features extracted from pre-trained networks remains challenging and relatively unexplored. For example, FID, KID, and IS use features extracted from an Inception model trained on ImageNet-1K [30]; LPIPS, and Intra-LPIPS, use features extracted from models trained on BAPPS [205] dataset. Although these pre-trained models effectively function as general-purpose feature extractors, their ability to capture properties/ attributes of out-of-domain data to objectively quantify the capabilities of GM-DC requires more investigation, *e.g.* medical images. In summary, the area of evaluation measures for GM-DC cannot be overstated, as it remains critical and challenging.

4 COMPREHENSIVE REVIEW

In this section, first, we will present our proposed taxonomy of approaches for GM-DC which systematically categorizes and organizes GM-DC methods under seven approaches (Tab. 3) based on the principal ideas of these methods. Then, we will discuss individual GM-DC methods organized under our proposed taxonomy.

Our Proposed Taxonomy of Approaches for GM-DC categorizes GM-DC methods into seven groups:

(1) **Transfer Learning:** In GM-DC, transfer learning (TL) aims to improve the learning of the generator for the target domain using the knowledge of a generator pre-trained on a source domain (with numerous and diverse samples). For example, some methods under this category use the knowledge of a StyleGAN2 pre-trained on the large FFHQ [77] to improve the learning of generation for face paintings by an artist, given only a few images of the artist's paintings [122, 189, 213]. Major challenges for TL-based GM-DC are to identity, select and preserve suitable knowledge of the source generator for the target generator. Along this line, there are six subcategories: i) *Regularization-based Fine-tuning*, explores regularizers to preserve suitable source generator's knowledge to improve learning target generator; ii) *Latent Space*, explores transformation/ manipulation of the source generator latent space; iii) *Modulation*, freezes and transfers weights of the source generator to the target generator and adds trainable modulation weights on top of frozen weights to increase the adaptation capability to the target domain; iv) *Natural Language-guided*, uses natural language prompt and supervision signal from language-vision models to adapt source generator to target domain; v) *Adaptation-Aware*, identifies and preserves the source generator's knowledge that is important to the adaptation task; vi) *Prompt Tuning*, is an emerging idea that freezes the weights of the source generator and learns to generate visual prompts (tokens) to guide generation for the target domain.

(2) **Data Augmentation:** Augmentation aims to improve GM-DC by increasing coverage of the data distribution with applying various transformations $\{T_k\}_{k=1}^{K}$ to available data. For example, within this category, some works augment the available limited data to train an unconditional StyleGAN2 [78] using the 100-shot Obama dataset or train a conditional BigGAN [12] with

only 10% of the CIFAR-100 dataset. A major challenge of these approaches is augmentation leakage, where the generator learns the augmented distribution, *e.g.* , generating rotated/ noisy samples. There are three representative categories: i) *Image-Level Augmentation*, applies the transformations on the image space; ii) *Feature-Level Augmentation*, applies the transformations on the feature space; iii) *Transformation-Driven Design*, leverages the information on each individual transformation T_k for an efficient learning mechanism.

(3) **Network Architectures:** These approaches design specific architectures for the generators to improve their learning under data constraints. Some works in this category design shallow/ sparse generators to prevent overfitting to training data due to over-parameterization. The primary challenge is that when endeavoring to design a new architecture, the process of discovering the optimal hyperparameters can be laborious. There are three major types of architectural designs for GM-DC: i) *Feature Enhancement*, introduces additional modules to enhance/ retain the knowledge within feature maps; ii) *Ensemble Large Pre-trained Vision Models*, leverages large pre-trained vision models to aid more accurate generative modeling, iii) *Dynamic Network Architecture*, evolves the architecture of the generative model during training to compensate for data constraints.

(4) **Multi-Task Objectives:** These approaches modify the learning objective of the generative model by introducing additional task(s) to extract generalizable representations and reduce overfitting under data constraints. As an example, some works define a pretext task based on contrastive learning [54] to pull the positive samples and push away negative ones in addition to the original generative learning task to prevent overfitting under limited available data. The efficient integration of the new objective with the generative learning objective could be challenging under data constraints. These works can be categorized into several kinds of approaches: i) *Regularizer*, adds an additional learning objective as a regularizer to prevent an undesirable behavior during training generative model under data constraint. Note that this category is different from regularizer-based fine-tuning as the latter aims to preserve source knowledge, but the former is for training without a source generator; ii) *Contrastive Learning*, adds the learning objective related to a pretext task to enhance the learning process of the generative model using additional supervision signal from solving this pretext task; iii) *Masking*, introduces alternative learning objective by masking a part of the image/ information to improve generative modeling by increasing the task hardness and preventing learning the trivial solutions; iv) *Knowledge Distillation*, introduces an additional learning objective that enforces the generator to follow a strong teacher; v) *Prototype Learning*, emphasizes learning prototypes for samples/ concepts within the distribution as an additional objective; vi) *Other Multi-Task Objectives*, include co-training, patch-level learning, and using diffusion to enhance generation.

(5) **Exploiting Frequency Components:** Deep generative models exhibit frequency bias tending to ignore high-frequency signals as they are hard to generate [153]. Data constraints can exacerbate this problem [197]. The approaches in this category aim to improve frequency awareness of the generative models by leveraging frequency components during training. For instance, certain approaches employ Haar Wavelet transform to extract high-frequency components from the samples. These frequency components are then fed into various layers using skip connections, aiming to alleviate the challenges associated with generating high-frequency details. Despite its effectiveness, utilizing frequency components for GM-DC has not been thoroughly investigated. The performance can be enhanced by incorporating more advanced techniques for extracting frequency components.

(6) **Meta-Learning:** These approaches create sample generation tasks with data constraints for the seen classes, and learn the meta-knowledge —knowledge that is shared between all tasks— across these tasks during meta-training. This meta-knowledge is then used in improving generative modeling for the unseen classes with data constraints. For instance, some studies, as meta-knowledge, learn to fuse the samples from the seen categories C_{seen} of the Flowers dataset [120] for sample generation. This meta-knowledge enables the model to generate new samples from unseen classes C_{unseen} within the same dataset ($C_{seen} \cap C_{unseen} = \emptyset$) by fusing only 3 samples from each class. Note that as these works employ episodic learning within a generative framework, the training stability can be impacted. Approaches within this line can be classified into three categories: i) *Optimization*, initializes the generative model with weights learned on the seen classes as meta-knowledge, to enable quick adaptation to unseen classes with limited steps of the optimization; ii) *Transformation*, learns cross-category transformations from the samples of the seen classes as meta-knowledge and applies them to available samples of the unseen classes to generate new samples; iii) *Fusion*, learns to fuse the samples of the seen classes as meta-knowledge, and applies learned meta-knowledge to sample generation by fusing samples of the unseen classes.

(7) **Modeling Internal Patch Distribution:** These approaches aim to learn the internal patch distribution within one image (in some cases a few images), and then generate diverse samples that carry the same visual content (patch distribution) with an arbitrary size, and aspect ratio. As an example, some works train a Diffusion Model using a single image of the "Marina Bay Sands", and after training, the Diffusion Model can generate similar images, but include additional towers topped by the similar "Sands Skypark". However, a major limitation of these methods lies in the fact that for every single image, usually a separate generative model is trained from scratch, neglecting the potential for efficient training through knowledge transfer in this context. Approaches proposed along this line can be categorized into two major groups: i) *Progressive Training*, progressively trains a generative model to learn the patch-distribution at different scales or noise levels; ii) *Non-Progressive Training*, learns a generative model at a single scale by implementing additional sampling techniques or new model architectures.

In what follows we delve into detailed descriptions of the approaches within each category.

4.1 Transfer Learning

Transfer Learning (TL) is a major approach for GM-DC. Given a source generator G_s (for GANs both G_s and D_s) pre-trained on a large and diverse source domain $\mathcal{D}_s$, these approaches aim to learn an adapted generator to the target domain G_t by initializing the weights to that of the source generator.

4.1.1 Regularizer-based Fine-Tuning. TGAN [182] is the first systematic study to evaluate transfer learning in GANs. TGAN shows that transfer learning reduces the convergence time and improves generative modeling under limited data. The knowledge transfer is performed by using the source GAN for initializing the weights of the target GAN, followed by fine-tuning the weights on target data. TGAN [182] demonstrates that: i) transferring D is more important than G, while transferring both G and D gives the best results; ii) transfer learning performance degrades by increasing the distance between source and target domains or decreasing the number of samples from target domain; iii) to select a pre-trained GAN for a target domain, in addition to a smaller distance, more dense source domains are preferable. As an example, for the Flower [120] target domain, surprisingly, a GAN pre-trained on semantically unrelated LSUN Bedrooms [201] is shown to be among the best sources [182].

W^3 [50] revisits the transfer learning in GANs with a modern structure —StyleGAN2-ADA [76, 78] instead of WGAN-GP [52] used in TGAN. Results in [50] suggest that for SOTA GANs, it is beneficial to transfer the knowledge from sparse and diverse sources (pre-trained StyleGAN2 on ImageNet) rather than dense and less diverse ones. One major limitation of TGAN is that under limited data simply fine-tuning the whole generator destroys a considerable portion of the general knowledge obtained on the source domain, and results in overfitting. Almost all of the following works aim to address this by different approaches to preserve the knowledge of the source generator.

BSA [121] limits scale and shift parameter updates during fine-tuning for batch normalization (BN) layers. FreezeD [112], hypothesizes that as D performs the discriminative task during training a GAN, based on common knowledge in discriminative learning [199], its early layers encode general knowledge which is shared between source and target domains. Therefore, this general knowledge is preserved during adaptation by freezing early layers of D. cGANTransfer [156] assumes that the pre-trained G is conditioned on class labels using BN parameters, *i.e.* each class has its own BN parameters [12]. Then, explicit knowledge propagation from seen classes to unseen classes is enforced by defining the BN parameters of the unseen classes to be the weighted average of the BN parameters of seen classes. SVD [139] applies singular value decomposition [173], and only updates the singular values that are related to changing entanglement between different attributes within data to constrain the parameter update.

EWC [103] aims to preserve the diversity of the source GAN during adapting to a target domain with only a few samples, *e.g.* , 10-shot. The importance of each parameter in source GAN is measured by Fisher Information (FI), and the change on each parameter during adaption is penalized based on its importance, *e.g.* , change over important parameters is penalized more. CDC [122] aims to keep the diversity of the generated samples using a cross-domain correspondence loss. Specifically, first, a batch of $N + 1$ latent codes are sampled for image generation: $\{G(z_0), ..., G(z_N)\}$. Then, using $G(z_0)$ as a reference, the similarity of generated samples to reference is measured for the generator before and after adaptation, resulting in two $N - way$ probability vectors. The diversity is preserved by adding the KL divergence between these two probability vectors to the standard loss as a regularizer. MaskD [220] applies random masks to extracted features of D, on top of CDC [122], to prevent overfitting. DDPM-PA [221] uses a pairwise adaptation method similar to CDC for adapting diffusion models to the new domain. RSSA [189] extends the cross-domain consistency idea of the CDC [122] to a more constrained form by preserving the structural similarity of the samples before and after adaptation. ProSC [114] extends RRSA by performing a progressive adaptation to the target domain in N iterations instead of a single adaptation to reduce the gap between pair of domains. CSR [49] uses a similar idea to CDC but applies semantic loss directly to the spatial space, *i.e.* , generated images with G_s and G_t.

C^3 [95], DCL [213], and CtlGAN [183] aim to preserve the diversity by applying contrastive learning. Assuming $G_s(z_i)$ as an anchor point, the generated sample for the same latent code with the adapted generator ($G_t(z_i)$) is considered a positive pair, and the generated samples with the adapted generator for other latent code values ($G_t(z_j)$, $i \neq j$) are considered as negative pairs. Additionally, DCL applies similar contrastive learning to the D.

JoJoGAN [23] addresses one-shot image generation using the style space of StyleGAN2. First, GAN inversion is used to find the corresponding style code of the reference image. Then, style mixing is used to generate a set of style codes, and generated images with these styles are used for GAN adaptation. GenOS [209] includes entity transfer with some related entity mask using an auxiliary network. D^3-TGAN [188] first inverts each target sample into the latent code space of the source GAN. Then, the maximum mean discrepancy between the features of the source G for inverted code and features of the adapted GAN for a random latent code is used as a regularizer.

FairTL [166] adopts transfer learning in GANs to train a fair generative model *w.r.t.* a sensitive attribute (SA) using a limited fair dataset. To model complex distributions like ImageNet, IC-GAN [14] learns data distribution as a mixture of conditional distributions. This enables IC-GAN to generate images from unseen distributions, by just changing the conditioning instances on the target samples. KDFSIG [69] exploits the knowledge distillation idea by treating the source model as a teacher and the target model as a student. F^3 [79] proposes a faster method to generate face images with features of a specific group. First, a GAN Inversion of target images is applied and then PCA is leveraged as a feature extraction strategy to render features of target group. DWSC [70] proposes the dynamic weighted semantic correspondence between the source and target generator during adaption to preserve the diversity.

4.1.2 Latent Space. MineGAN [180] trains a miner network M during adaptation, to map the latent space z of the source GAN to another space $u = M(z)$ more appropriate for the target domain. MineGAN++[181] extends MineGAN by only updating important parameters. GenDA [193] proposes a lightweight attribute adaptor in the form of scaling and shifting latent codes to adapt the latent space of the source GAN to the target GAN. LCL [113] freezes G and learns a network to map the latent codes from the $\mathcal{Z}$ space to the extended intermediate space $\mathcal{W}^+$ of a pre-trained StyleGAN2 during adapting GAN. WeditGAN [35] proposes to learn a constant offset parameter (Δw) for the target domain in the intermediate latent space of StyleGAN2 to relocate source latent codes to the target domain. After fine-tuning the generator to a target domain, SiSTA [167] perturbs latent representations of the fine-tuned generator that falls below a threshold, either by replacing them with zero or reverting them back to the pre-trained generator's weights. MultiDiffusion [8] freezes the whole parameters of the source diffusion model and optimizes the latent code as a post-processing method to generate the desired output based on a conditioned input.

4.1.3 Modulation. In signal processing literature, modulation varies some key attributes of a signal to add the desired information to it [123]. Similarly, in deep neural networks, modulation is used to add some desired information to a base network by adding modulation parameters to the parameters/ features of the base network. AdaFMGAN [210] shows that layers closer to the sample (earlier layers in D, and later layers in G) encoder general knowledge. This general knowledge is conceptually shared between source and target domains and aimed to be preserved by Adaptive Filter Modulation which trains a scale and shift parameter for each $k \times k$ kernel. GAN-Memory [26], and CAM-GAN [174] use similar modulation ideas to modulate a pre-trained GAN for generative continual learning. AdAM [212] uses kernel modulation [1] for few-shot generative modeling by aiming to preserve the important wights of a pre-trained GAN during a few-shot adaptation to a target domain. HyperDomainNet [5] adds an additional modulation to StyleGAN2 [78] for adapting to a new domain, while optimizing only $6K$ parameters.

4.1.4 Adaptation-Aware. Adaptation-aware transfer learning approaches propose that different parts of the knowledge encoded on a pre-trained generative model could be important based on the target domain. AdAM [212] proposes a probing step before the main adaptation, where the importance of each kernel for adapting a source GAN to the target domain using a few samples is measured using FI. Then, during the main adaptation, the important kernels are preserved using modulation, and the other kernels are simply fine-tuned. RICK [214] shows that incompatible knowledge from a source domain to a target domain is related to the kernels with the least importance to this adaptation, and this knowledge can not be removed by simple fine-tuning. Therefore, RICK proposes a dynamic probing schedule during adaptation where it gradually prunes the kernels with the least importance.

4.1.5 Natural Language-Guided. Vision-Language models like CLIP [132] are usually trained on large-scale image-text pairs and learn to encapsulate the generic information by combining image and text modalities. This generic information is shown to be helpful in various downstream tasks including zero-shot and few-shot image generation. StyleGAN-NADA (NADA) [44] is the pioneering work on utilizing CLIP for zero-shot image generation. NADA [44] uses a text prompt T_t –which describes the target domain– as input and uses feedback to adapt a pre-trained StyleGAN2 to the target domain. Specifically, assuming a text prompt T_s that describes the source domain (*e.g.* , "Photo" for a StyleGAN2 pre-trained on the FFHQ), and a given T_t (*e.g.* , "Fernando Botero Painting"), CLIP's text encoder E_T is used to find the update direction in the embedding space: $\Delta T = E_T(T_t) - E_T(T_s)$. Similarly, the direction of the update/ change for the images can be computed using generated images with source and target generators: $\Delta I = E_I(G_t(z)) - E_I(G_s(z))$, where E_I denotes CLIP's image encoder. Since the image and corresponding texts are aligned in the CLIP space, NADA [44] proposes to update the generator's parameters in a way to match ΔI and ΔT leading to the directional loss $\mathcal{L}_{directional}$:

$$\mathcal{L}_{directional} = 1 - \frac{\Delta I(G_s(z), G_t(z)).\Delta T(T_s, T_t)}{|\Delta I(G_s(z), G_t(z))||\Delta T(T_s, T_t)|} \tag{7}$$

This idea can be easily extended to one-shot image generation, by replacing ΔT with the direction obtained by target image I_t and a batch of generated images by the source generator: $\Delta I' = E_I(I_t) - \mathbb{E}_i\{E_I(G_s(z_i))\}$, where $\mathbb{E}_i\{E_I(G_s(z_i))\}$ denotes the mean of the CLIP embedding for a batch of generated images. MTG [224] extends the idea for one-shot image generation by replacing the mean embedding with the projection of the target image on the source generator denoted as I_s^*. Specifically, MTG uses GAN inversion to get the corresponding z^{ref} for I_t, and uses it to generate the projected image: $I_s^* = G_s(z^{ref})$. HyperDomainNet improves the performance of the NADA and MTG by freezing the weights of the source generator and training modulation weights for the synthesis part inside the generator. DiFa [207] adds an attentive style loss to directional loss of NADA [44] as a local-level adaptation which aligns the intermediate tokens of the generated image with source and pre-trained GAN. OneCLIP [93] exploits the CLIP embedding for three major modules in one-shot learning: i) inverting sample into latent space, ii) preserving the diversity of the GAN during adaptation, and iii) a patch-wise contrastive learning approach for preserving local consistency.

IPL [53] mentions that NADA [44] uses a fixed update direction for a target domain due to manual text prompts being used for describing source and target domains. To address this, [53] learns a specific prompt for each generated image, *e.g.* , "*Elder glass man* photo" → "*Elder glass man* Fernando Botero Painting". A mapper function F is used to learn prompts for the generated images $G_s(z_i)$:

$$F(z_i) = [V]_1^i[V]_2^i...[V]_m^i \tag{8}$$

where $[V]_j^i$ represents the j^{th} prompt vector. Then, the domain embedding $[Y_s]$ is added to this prompt to represent both the prompt and the domain: $M_s^i = F(z_i)[Y_s]$. The mapper function is trained by contrastive loss in the CLIP space. After training F, the same adaptation process as NADA [44] can be used but using $T_s(z_i) = [V]_1^i[V]_2^i...[V]_m^i[Y_s]$.

DreamBooth [143] addresses subject-driven sample generation by fine-tuning a text-to-image diffusion model e.g. Imagen [144], DALL-E2 [134]. Input images are paired with a text prompt that contains a unique identifier and the subject class (e.g., "A [V] dog"), and the pair is used to fine-tune the model. They further propose a class-specific prior preservation regularizer to encourage diversity and to mitigate *language drift*, i.e., the model progressively loses syntactic and semantic knowledge during fine-tuning. Textual-Inversion [43] optimizes a word vector for the new subject given a few images and uses that word vector for the subject-driven generation. MCC [91]

extends DreamBooth by the capability of adding multiple subjects and improves convergence and performance by restricting fine-tuning to a subset of cross-attention layer parameters in DM. BLIP-Diffusion [96] uses BLIP-2 [97] multimodal encoder to extract a more text-aligned representation for each subject, and then a subject representation learning step is performed to enable DM to leverage such representation for fast and high-fidelity subject-driven sample generation. SINE [208] uses a similar fine-tuning of text-to-image diffusion model with a unique identifier, but at the patch level. In addition, the mixing of the latent code is also used to edit the subject and put it in a new context. SpecialistDiffusion [109] addresses adapting the text-to-image models to an unseen style given a few samples from that style using augmenting both text and image, and sparse diffusion for computation efficiency.

4.1.6 Prompt Tuning. VQ-VAEs (Sec. 2.1) can be broadly categorized into two types from the perspective of predicting the latent prior of visual tokens. AutoRegressive (AR) approaches like DALL·E [135] and VQ-GAN [38], learn an AR predictor that follows a raster scan order and predicts the visual tokens from left to right, line-by-line. Non-AutoRegressive (NAR) approaches like DALL·E2 [134], MaskGIT [19], Latent Diffusion [140], or Imagen [144] usually resort to masking techniques [31] to predict the visual tokens in a series of refinement or denoising steps. VPT [160] is the first work that adopts the prompt tuning idea for image generation with generative knowledge transfer. It uses a VQ-VAE framework where a MaskGIT[19]/ VQ-GAN[38] on the ImageNet dataset (as an example of NAR/ AR approach) is used as a pre-trained network. Then, during adaptation, all the parameters of the VQ Encoder, VQ decoder, and transformer are frozen, and a generator is learned to minimize the adaptation loss by generating and appending a set of visual tokens to the predicted prior. These visual tokens guide the generation process for the target domain by helping the transformer to predict proper tokens to the VQ decoder.

4.2 Data Augmentation

Data augmentation increases the quantity and diversity of the training data which is shown to be beneficial for GM-DC. However, if it is not deployed correctly, augmentations can leak into the generator resulting in generating samples with similar augmentations, *e.g.* noisy or rotated images, which is undesirable.

4.2.1 Image-Level Augmentation. CR-GAN [204] and bCR [216] apply various transformations on images and enforce the output of the generator to be the same for original and transformed images. Even though not developed for GM-DC, experimental results in [76] show that CR-GAN and bCR are beneficial for limited data scenarios. ADA proposes applying the transformations to real and fake images but with a probability $p < 1$. The central design in ADA [76] is that the strength of the augmentation (p) is being adapted based on the training dynamics. Specifically, the portion of the real images that get positive output from the discriminator, *i.e.* , $r = \mathbb{E}[Sign(D)]$, is used as an indicator of the discriminator overfitting ($r = 0$ no overfitting, and $r = 1$ complete overfitting). Then, during training, p is adjusted to keep r low. DiffAugment [211], and IAG [217] use a similar idea to ADA, but without the adaptive component ($p = 1$). APA [74] uses the same adaptive augmentation mechanism in ADA, but instead of using transformations like rotation, it randomly labels generated images as pseudo-real ones to prevent an overconfident discriminator.

DiffusionGAN [185] applies the gradual diffusion process on real and generated images during training GAN. Training starts with real and generated images, and each diffusion step is applied after a certain number of training epochs making the bi-classification task harder for the discriminator to prevent its overfitting. PatchDiffusion [184] augments the data during training diffusion models by sampling patches with random locations and random sizes alongside the full image and conditioning the denoising score function [75] on the patch size and the location information.

4.2.2 Feature-Level Augmentation. AdvAug [21] computes the adversarial perturbation δ for the feature maps of the discriminator and generator using the projected gradient descent [110]. Denoting the discriminator as $D = D_2 \circ D_1$, the adversarial augmentation is applied on the intermediate feature maps (D_1) of both real and generated images, resulting in $D_1(x) + \delta$, and $D_1(G(z)) + \delta$. The adversarial loss is then added to the loss function of D during GAN training to maximize the score of the perturbed real image and minimize the score of the perturbed generated image:

$$\mathcal{L}_D^{adv} := \max_{\|\delta\|_\infty < \epsilon} \mathbb{E}_{x \sim p_{data}}[f_D(-D_2(D_1(x) + \delta))] + \max_{\|\hat{\delta}\|_\infty < \epsilon} \mathbb{E}_{z \sim p_z}[f_D(D_2(D_1(G(z)) + \hat{\delta}))] \quad (9)$$

As AdvAug is performed on the feature level, it is shown to be complementary to image-level augmentations like ADA [76] and DiffAug [211]. AFI [29] observes a flattening effect in discriminators with multiple output neurons, and takes advantage of this observation by proposing feature interpolation as implicit data augmentation.

4.2.3 Transformation-Driven Design. DAG [168] uses a separate discriminator D_k for discriminating real and fake images that are augmented by transformation T_k. A weight-sharing mechanism between all discriminators is used to prevent overfitting. Additionally, DAG provides a theoretical ground for training convergence under augmentation. As mentioned in [48], for an optimal discriminator D^*, optimizing G is equivalent to minimizing the Jensen-Shannon (JS) divergence between the real data distribution P_{data} and generated data distribution P_{model}, *i.e.* , $JS(P_{data}, P_{model})$. Denoting P^T_{data}, and P^T_{model} as the distribution of the real and generated data under augmentation T, [168] shows that JS divergence between two distributions is invariant under differentiable and invertible transformations:

$$JS(P_{data}, P_g) = JS(P^T_{data}, P^T_g) \quad (10)$$

This means that as long as the augmentation is differentiable and invertible, training convergence is guaranteed. SSGAN-LA [68] extends DAG by merging all discriminators to a single discriminator and augmenting the label space of the discriminator, *i.e.* , asking D to detect the type of augmentation in addition to conventional real/ fake detection.

4.3 Network Architectures

4.3.1 Feature Enhancement. FastGAN [107] proposes a light-weight GAN structure —shallower G and D compared to SOTA GANs like StyleGAN2— to decrease the risk of overfitting. Inspired by skip connections [55], and squeeze-and-excitation module [71], FastGAN fuses features with different resolutions in G through proposed skip-layer excitation modules. An additional reconstruction task is defined for D. MoCA [100] learns some prototypes for each semantic concept within a domain, *e.g.* , railroad, or sky in a photo of a train. Then, by attending to these prototypes during image generation, some residual feature maps are produced to improve image generation. DFSGAN [194] proposes to preserve the content and layout information in intermediate layers of G by extracting channel-wise and pixel-wise information and using them to scale corresponding feature maps. SCHA-VAE [46] extend latent variable models for sets to a fully hierarchical approach and propose Set-Context-Hierarchical-Aggregation VAE for few-shot generation.

4.3.2 Ensemble Pre-trained Vision Models. ProjectedGAN [149] proposes to project real and generated images into the feature space of a pre-trained vision model to enhance D's performance in discriminating real and fake images by adding two modules. First, the output from multiple layers is used with separate discriminators. Then a *random projection* is used to dilute the features and encourage the discriminator to focus on a subset of the features. Vision-aided GAN [92] uses an ensemble of the original discriminator D and additional discriminators $\{\hat{D}_n\}_{n=1}^N$ to perform the classification task. The additional discriminators $\{\hat{D}_n\}_{n=1}^N$ have a set of pre-trained feature

extractors $\mathcal{F} = \{F_n\}_{n=1}^{N}$ (extracted form pre-trained vision models) with a small trainable head C_n added on top: $\hat{D}_n = F_n \circ C_n$.

4.3.3 Dynamic Network Architecture. CbC [155] shows that under data constraints, where an unconditional GAN can generate satisfactory performance, training the conditional GANs (cGANs) result in mode collapse. To mitigate this issue, CbC [155] starts training from an unconditional GAN and slowly transitions to a cGAN using a transition function $0 \leq \lambda_t \leq 1$. Considering the conditioning variable as c, this transition is implemented in G as $G(z, c, \lambda_t) = G(S(z) + \lambda_t.E(c))$, with S and E as neural networks that transform the latent code and the conditioning variable. DynamicD [191] dynamically reduces the capacity of D by randomly sampling a subset of channels of D during each training iteration to prevent overfitting. Inspired by the lottery ticket hypothesis [41], AdvAug [21] and Re-GAN [151] have shown that a much sparse subnetwork of the original generator can be useful for GM-DC. AutoInfoGAN [157] applies a reinforcement learning-based neural architecture search to find the best network architecture for the generator.

4.4 Multi-Task Objectives

4.4.1 Regularizer. LeCam [169] uses two moving average values to track D's prediction for real and generated images, denoted by α_R and α_F, respectively. Then the distance between the D's prediction for real (fake) images and α_F (α_R) is decreased by adding a regularizer to prevent overfitting. Analysis in [169] shows that this regularizer enforces WGAN [7]/ BigGAN[12] to minimize the LeCam-divergence which is beneficial for GM-DC. Reg-LA [67] uses a similar idea to regularize the label-augmented GANs discussed in Sec. 4.2. DIG [39] shows that the discriminator gradient gap between real and generated images increases when training GANs with limited data, and adds this gap as a regularizer to prevent this behavior. MDL [87] addresses the pre-training free few-shot image generation by adding a regularizer that aims to keep the similarities between the latent codes in $\mathcal{Z}$ space and corresponding generated images in image space.

4.4.2 Contrastive Learning. InsGen [192] uses contrastive learning to improve learning D by introducing a pretext task. The pretext task is defined as instance discrimination, meaning that each sample should be mapped to a separate class. This is done by constructing the query and key from the same sample as positive pair, and all remaining images as negative pair. FakeCLR [104] analyze three different contrastive learning strategies, namely instance-real, instance-fake, and instance-perturbation. It is shown that instance-perturbation contributes the most improvement in quality and can effectively alleviate the issue of latent space discontinuity. Note that contrastive learning is also combined with other approaches as discussed before (C^3 [95], DCL [213], CtlGAN [183], CML-GAN [128], and IAG [217]).

4.4.3 Masking. MaskedGAN [72] utilizes a masking idea for training GANs under limited data by masking both spatial and spectral information. For spatial masking, they use a patch-based mask to enable random masking of all spatial parts. For spectral masking, they mask each frequency channel (extracted by Fourier transform) based on the amount of information, *i.e.* , channels with more information are more probable to be masked. MaskD [220] randomly masking feature maps extracted by D for a few-shot setup. DMD [206] detects that the discriminator slows down learning and applies random masking to its features adaptively to balance its learning pace with the generator.

4.4.4 Knowledge Distillation. KD-DLGAN [28] proposes a knowledge distillation (KD) [18, 58] approach by leveraging CLIP [132] as the teacher model to distill text-image knowledge to the discriminator. They propose two designs: aggregated generative knowledge designs a harder

learning task, and correlated generative knowledge distillation improves the generation diversity by distilling and preserving the diverse image-text correlation from CLIP. As discussed before, KDFSIG [69] also uses KD in the context of transfer learning for few-shot image generation.

4.4.5 Prototype Learning. Inspired by the recent success of learning prototypes in few-shot classification, ProtoGAN [195], aims to improve the fidelity and diversity of the FastGAN under limited data[159]. ProtoGAN has two main modules: prototype alignment for increasing the fidelity of the generated images, and diversity loss to improve the generation diversity. MoCA [100] also learns prototypes but for different semantic concepts through an attend and replace mechanism on the extracted feature maps of G.

4.4.6 Other Multi-Task Objectives. Gen-Co [27] uses multiple discriminators to extract diverse and complementary information from samples. This *co-training* has two major modules: weight-discrepancy co-training which trains separate Ds with different weights, and data-discrepancy co-training which in addition to training separate Ds also uses different information as inputs, *i.e.* , spatial or frequency information. AdaptiveIMLE [3] proposes an adaptive version of implicit maximum likelihood estimation [98] to improve the mode coverage by assigning different boundary radii for each sample. PathcDiffusion [184] and AnyResGAN [15] show the effectiveness of *Patch-Level* learning of the generators. Diffusion-GAN [185] leverages the *diffusion process* to improve training GANs by gradually increasing the task hardness for D. D2C [158] uses a DM to improve the sampling process of VAEs by denoising the latent codes and feeding VAE with a clean latent code for sample generation. FSDM [45] uses an attentive conditioning mechanism and aggregates image patch information using a vision transformer for image generation for unseen classes.

4.5 Exploiting Frequency Components

Approaches in this category aim to improve frequency awareness to improve GM-DC. FreGAN [197] extracts high-frequency information (HF) of images (related to details in images) using Haar Wavelet transform [130] and uses three different modules to emphasize learning high-frequency information: high-frequency discriminator uses HF as an additional signal to perform real/fake classification, frequency skip connection feeds the HF information of each feature map to the next one in G to prevent frequency loss, and a frequency alignment loss is used to make sure G and D are learning frequency information in the same pace. WaveGAN [196] uses similar idea, but in a different setup to address task cGM-2. Gen-Co extracts some frequency information of the image and feeds it to a separate D in addition to using original real and fake images. MaskedGAN [72] masks out some frequency bands of the input during training to enforce the generative model to focus more on under-represented frequency bands.

4.6 Meta-Learning

Meta-learning shifts the learning paradigm from data level to task level to capture across-task knowledge as *meta-knowledge*, and then adapt this meta-knowledge to improve the learning process of unseen tasks in the future. An abundant of recent works adopt meta-learning to tackle few-shot classification [1, 40, 159, 162, 175] and few-shot semantic segmentation [177]. These works usually follow the *episodic learning* setup which matches the way that model is trained and tested. Considering task distribution $P_{\mathcal{T}}$, a set of training tasks are constructed from seen classes $\mathcal{T}^{train} = \{\mathcal{T}_i^{train}\}$, where $\mathcal{T}_i^{train}$ denotes i^{th} training (meta-training) task. The model is trained on the meta-training tasks and later tested on the meta-test tasks $\mathcal{T}^{test} = \{\mathcal{T}_j^{test}\}$ constructed from unseen classes. Usually meta-training and meta-testing tasks follow the same distribution $P_{\mathcal{T}}$. Similarly, the approaches in this category use meta-learning to address image generation: train a generative model on a set of few-shot image generation tasks constructed from seen classes of

a domain, then, test it on the few-shot image generation tasks from unseen classes of the same domain.

4.6.1 Optimization. Optimization-based meta-learning algorithms are used in these approaches for learning meta-knowledge. Generative Matching Network (GMN) proposes a similar attention mechanism used in Matching Networks [175] for few-shot image generation with variational inference. FIGR [25] meta-trains a GAN using Reptile [117]. Training has an inner loop that adapts the GAN weights based on a few-shot image generation task and an outer loop that updates the meta-knowledge using Reptile. Dawson [106] modifies the inner loop training to directly get the gradients for the generator from evaluation data. FAML [127] uses a similar idea to FIGR, but instead of using the standard GAN structure, it uses an encoder-decoder architecture for the generator. CML-GAN [128] extends FAML [127] by leveraging contrastive learning to learn quality representations.

4.6.2 Fusion. MatchingGAN [62] learns to generate new images for a category by fusing the available images of that category. A set of encoders are used to estimate the similarity between the embedding of the latent code and input images. Then, these similarities are used as interpolation coefficients by an auto-encoder to extract the embeddings of the training images and fuse them for generating new images. F2GAN [65] uses random coefficients for general information, and attention module for details. The attention module takes the weighted average of the real image features and the corresponding features from the decoder to produce the image details. LoFGAN [51] focuses on local features in the fusion process. Given a batch of images, one sample is selected as a base while the remaining are utilized as a reference set. This set acts as a feature bank for the fusing process. WaveGAN [196] adds frequency awareness to LofGAN by extracting and feeding the frequency components of feature maps to later layers of the generator. AMMGAN [101] utilizes an adaptive fusion mechanism for learning pixel-wise metric coefficients during the fusion.

4.6.3 Transformation. DAGAN [6] leverages the task of the learning augmentation manifold in the learning process of the GAN. This is modeled as some transformations on the input, and these transformations are applied to the new sample from the unseen classes for sample generation. DeltaGAN [63] learns the difference between images (delta) in the feature space, and then uses this delta concept for diverse sample generation. Disco [64] learns a dictionary based on seen images to encode input images into visual tokens. These tokens are then fed into the decoder with the style embedding of seen images to generate images from unseen classes. AGE [34] uses GAN inversion to invert the samples of a category to $\mathcal{W}^+$ space of StyleGAN2 [78]. The mean latent code for all samples of a category is used as a prototype and all differences are considered as general attributes. These attributes are then used to diversify sample generation for unseen classes. SAGE [33] addresses the class inconsistency in AGE by taking all given samples from unseen classes into account during inference. HAE [99] uses a similar idea to AGE [34], but in the Hyperbolic space instead of using Euclidian distance which allows more semantic diversity control. LSO [218] finds a prototype for each class similar to AGE [34]. Then it adjusts GAN to produce similar images to target samples using latent samples from the neighborhood of the prototype, followed by updating the prototype in latent space using the adapted GAN.

4.7 Modeling Internal Patch Distribution

4.7.1 Progressive Training. SinGAN [154] is the pioneering work that makes use of the internal distribution of the patches within an image to train a generative model. It trains a pyramid of generators $\{G_0, \ldots, G_N\}$ against a pyramid of real images $\{x_0, \ldots, x_N\}$, where x_n is a downsampled

version of input image x by a factor of r^n. The generator at scale n uses random noise z_n and upsampled version of the generated image from the lower resolution $\tilde{x}_{n+1}$ as input: $\tilde{x}_n = G_n(z_n, (\tilde{x}_{n+1}) \uparrow^{r_n})$. Similarly, a pyramid of discriminators is used where D_n compares the $\tilde{x}_n$ and x_n in patch-level for real-fake classification. SinDDM [89] applies the same idea but uses diffusion models with a fully convolutional lightweight denoiser. ConSinGAN [59] stacks the new layers for a bigger scale on top of the previous layers used for a smaller scale instead of using separate generators for each scale. BlendGAN [85] and DEff-GAN [90] extend previous approaches for learning the internal distribution for k images, thereby allowing for the potential mixing of different image semantics and improving diversity. SinDiffusion [178] addresses artifacts in SinGAN due to progressive resolution growth by applying progressive denoising using a diffusion model architecture.

4.7.2 Non-Progressive Training. One-Shot GAN [163] uses a standard generator (single-scale), but multiple paths for the discriminator to enforce learning objects' appearance and how to combine them. Within the discriminator the low-level loss is defined on low-level features and two different losses are defined to learn the content and the layout in image patches. SinFusion [119] explores learning the internal patch distribution from both a single image and video. SinFusion extends on DPPM [61] and reduces the size of the receptive fields by first removing attention layers, then adopting ConvNext [108] blocks in the U-Net [141] architecture. To reconstruct videos, a series of images are fed into a series of 3 identical models. The first model predicts the next frame; the second model denoises and removes small artifacts from the generated images; the last model interpolates between the different frames.

5 DISCUSSION

Here, we present an analysis of the literature and discuss the research gap and future directions in GM-DC.

5.1 Analysis of the Research Landscape

In this work, we propose a **taxonomy of eight different tasks for GM-DC** (Fig. 1, Tab. 2) based on the problem setups of GM-DC publications. Our investigation of the literature focusing on each task (Fig. 3) reveals that a significant portion of the works (up to 80%) concentrate on unconditional generation, either through training from scratch or adapting from a pre-trained model. Additionally, zero-shot unconditional generation is beginning to attract more attention. Similarly, adaptation for in-domain classes has garnered considerable interest for conditional generation. Meanwhile, conditional generation for out-of-domain classes via adaptation has not been explored adequately. Furthermore, subject-driven generation, which enables more control over content generation, is an emerging task. We anticipate increasing interest on this task as recent text-to-image generative models become more accessible.

We further present a **taxonomy of approaches for GM-DC** (Fig. 1, Tab. 3) as our another contribution. Our study reveals that transfer learning is a predominant solution for GM-DC, capable of tackling a large number of tasks (specifically, 5 out of 8 tasks, as indicated in Tab. 3 and Fig. 1), while effectively handling all data constraints including limited data, few-shot, and zero-shot. Moreover, ≈39% of the studies propose new methods based on transfer learning (Fig. 3). More than 20% of the studies propose methods based on other approaches that are compatible to transfer learning, *e.g.* data augmentation. These methods could be used with transfer learning-based methods to improve performance. The primary challenges in transfer learning are selection and preservation of source knowledge useful for generating high-quality and diverse target domain samples. Adaptation-aware approach [212, 214] could be a sound direction in this aspect where they consider both source and target domains (the adaptation process) for knowledge preservation.

Language-guided approaches [5, 44, 53, 224] are gaining increasing attention due to their ability to facilitate zero-shot generation through appropriate application of vision-language models during the transfer learning phase. Visual prompt tuning [160] is a recent method, which guides the generation of target domain samples by generating visual tokens.

Data augmentation [76, 168, 211] remains a potent technique in GM-DC where it boosts performance under limited data by increasing coverage of the data distribution through various transformations. Multi-task objectives [72, 169, 192] which incorporate additional learning objectives are usually complementary to data augmentation. Various network architecture designs [100, 107] that aim to prevent overfitting or preserve the feature maps are also shown to be significantly effective for GM-DC. Given that generative models tend to exhibit biases in capturing frequency components, enhancing the frequency awareness in these models is an emerging direction for GM-DC [197]. Meta-learning [25] enables generative models to learn inter-task knowledge from seen classes, and then handle new generation tasks from unseen classes usually without fine-tuning [51, 63]. Internal patch-distribution modeling [119, 154] effectively trains a generative model from scratch using a single reference image (scene) to produce novel scene compositions.

Regarding the types of generating models, our study shows that around 86% of the GM-DC works focus on GANs (Fig. 3). This preference can be attributed to the extensive research in GANs. Recently, there has been a growing interest in DMs (12%) and VAEs (3%), particularly VQ-VAE, driven by the success of DMs [134, 144] and transformer-based token prediction methods in generative modeling [19, 38]. We anticipate increasing attention directed toward DMs and VQ-VAEs. Furthermore, our survey reveals an interesting trend: around 64% of the works focus on addressing the challenging task of few-shot learning, while 33% concentrate on limited data scenarios. While only 3% of works address zero-shot learning, we expect growing interest due to recent advancements in vision-language models [93, 102].

5.2 Research Gap and Future Directions

5.2.1 Harnessing the power of foundation models. As previously discussed, transfer learning is a prominent and highly effective solution for GM-DC. Nevertheless, the majority of existing literature uses pre-trained StyleGAN2 (FFHQ) or BigGAN (ImageNet) networks as source models. A potential future direction for GM-DC is to explore the capabilities of foundation models [11], *i.e.* large models trained using massive amounts of data. In particular, recent text-image generation models including DALL·E-2 [134] (≈3.5B parameters), Imagen [144] (≈4.6B parameters) and Stable Diffusion [140] (≈890M parameters) encode knowledge regarding a wide range of concepts for high-quality, diverse image generation. Leveraging such foundation models for GM-DC is relatively under-explored.

5.2.2 Grounding zero-shot image generative capabilities. Recent studies have demonstrated the feasibility of zero-shot image generation for well-known concepts, *e.g.* "Tolkien Elf" [44]. However, grounding zero-shot image generation models to generate evolving/ new semantic concepts remains a relatively unexplored and challenging area. For instance, how to generate an image depicting "The coronation of Charles III and Camilla as King and Queen of the United Kingdom," an event that occurred in May 2023, that related images may not be captured by existing models. This requires strategies that allow continual learning, semantic concept editing, and the incorporation of temporal contexts.

5.2.3 Knowledge transfer for distant/ remote target domains. Knowledge transfer has received significant attention in GM-DC research. Many works concentrate on utilizing pre-trained knowledge of a source domain to enhance learning in the target domain, as evident from the statistics in Fig. 1 and Fig. 3. However, we remark that exploring knowledge transfer for modeling target domains which are distant/ remote from the source domains still remains largely unexplored. This problem

is challenging, as demonstrated in our experiment to transfer knowledge from Human Faces → Flowers (Fig. 6), which clearly demonstrates the complexity of the task. We urge more investigation in knowledge transfer for modeling distant/ remote target domains in GM-DC research.

5.2.4 Exploring different types of generative models for GM-DC. Our analysis reveals that around 85% of GM-DC works focus on GANs and there is less attention to other types of generative models. Meanwhile, recent generative models such as diffusion models have made a lot of progress, achieving comparable performance to GANs in terms of quality and diversity of generated samples [32]. We remark that recent generative models are fundamentally different from GANs, *e.g.* multiple iterations are required to generate samples in diffusion models, suggesting that current GM-DC methods developed for GANs could be sub-optimal for other types of generative models.

5.2.5 Holistic evaluation of GM-DC. Evaluation of GM-DC presents multiple challenges including difficulties in estimating real data statistics under low-data regimes, lack of unified framework for human evaluation of GM-DC samples, and heavy reliance on particular (pre-trained) feature extractors to quantify the capabilities of GM-DC. In particular, developing holistic evaluation frameworks integrating both objective measurements and subjective judgements tailored for different tasks is essential for understanding GM-DC capabilities. Advancing holistic evaluation is important for GM-DC methods to be applied in a variety of real-world scenarios.

5.2.6 Data-centric approaches for GM-DC. We remark that data-centric approaches [187] for advancing GM-DC have been relatively overlooked in the literature. Majority of GM-DC methods focus on advancing training procedures based on a given set of training samples, but little attention has been put on how GM-DC performance may be affected by characteristics of the given training samples. Particularly, for GM-DC problems, where a domain is described using limited training samples, the characteristics of the samples can have noticeable impact on performance of GM-DC methods, as hinted in our analysis (see Fig. 7). We suggest greater emphasis on data collection, curation and pre-processing for GM-DC advancement.

5.3 Beyond Image Generation

Existing GM-DC works focus on image generation primarily. There are a few recent works to study other data types. [222] studies *3D shape generation* under few-shot target data (10-shot) utilizing pre-trained 3D generative models and optimization adaptation to retain the probability distributions of pairwise adapted samples. CLIP-Sculptor [146] leverages CLIP guidance for zero-shot *shape generation*. [176] studies few-shot *font generation* which aims to transfer the source domain style to the target domain. In particular, they introduce a content fusion module and a projected character loss to improve the quality of skeleton transfer for few-shot font generation. [13] explores the problem of few-shot *semantic image generation* where the objective is to generate realistic images based on semantic segmentation maps. Their approach employs transfer learning on both GANs and DMs for few-shot semantic image synthesis.

6 CONCLUSION

Generative Modeling under Data Constraints (GM-DC) is a burgeoning research area. This survey delves into this field by meticulously examining research papers in this area, encompassing different types of generative models including VAEs, GANs, and Diffusion Models. Drawing from this analysis, we identify several challenges encountered in GM-DC, including those related to training, data selection, and model evaluation. Moreover, we propose two taxonomies to categorize works related to GM-DC: a task taxonomy that identifies the variety of generation tasks, and an approach taxonomy that categorizes the extensive list of solutions for these tasks. We present a Sankey

diagram to illuminate the interactions between different GM-DC tasks, approaches, and methods. Additionally, we present an organized review of existing GM-DC works and discuss research gaps and future research directions. Our aspiration is that this survey not only could offer valuable insights to researchers but also help spark further advancements in GM-DC.

Ethics Statement. Generative models could be mis-used to disseminate mis- and disinformation due to their ability to generate realistic content. In particular, advanced generative models could be mis-used by malicious users to fabricate deepfake images, portraying individuals engaging in actions they never actually performed. Advances in GM-DC could exacerbate the situation as it becomes possible to generate realistic content with less data. We advocate for ethical and responsible usage of GM-DC methods and studying of mitigation techniques [16, 17, 111, 186, 215].

REFERENCES

[1] Milad Abdollahzadeh, Touba Malekzadeh, and Ngai-Man Man Cheung. 2021. Revisit multimodal meta-learning through the lens of multi-task learning. In *Advances in Neural Information Processing Systems*.

[2] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. 1985. A learning algorithm for Boltzmann machines. *Cognitive science* 9, 1 (1985), 147–169.

[3] Mehran Aghabozorgi, Shichong Peng, and Ke Li. 2023. Adaptive IMLE for few-shot pretraining-free generative modelling. In *Proceedings of the International Conference on Machine Learning*.

[4] Alper Aksac, Douglas J Demetrick, Tansel Ozyer, and Reda Alhajj. 2019. BreCaHAD: A dataset for breast cancer histopathological annotation and diagnosis. *BMC research notes* 12, 1 (2019), 1–3.

[5] Aibek Alanov, Vadim Titov, and Dmitry P Vetrov. 2022. HyperDomainNet: Universal domain adaptation for generative adversarial networks. In *Advances in Neural Information Processing Systems*.

[6] Antreas Antoniou, Amos Storkey, and Harrison Edwards. 2017. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340* (2017).

[7] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein generative adversarial networks. In *Proceedings of the International Conference on Machine Learning*.

[8] Omer Bar-Tal, Lior Yariv, Yaron Lipman, and Tali Dekel. 2023. Multidiffusion: Fusing diffusion paths for controlled image generation. In *Proceedings of the International Conference on Machine Learning*.

[9] Sergey Bartunov and Dmitry Vetrov. 2018. Few-shot generative modeling with generative matching networks. In *International Conference on Artificial Intelligence and Statistics*.

[10] Mikołaj Bińkowski, Dougal J. Sutherland, Michael Arbel, and Arthur Gretton. 2018. Demystifying MMD GANs. In *International Conference on Learning Representations*.

[11] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[12] Andrew Brock, Jeff Donahue, and Karen Simonyan. 2019. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*.

[13] Marlène Careil, Jakob Verbeek, and Stéphane Lathuilière. 2023. Few-shot semantic image synthesis with class affinity transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[14] Arantxa Casanova, Marlene Careil, Jakob Verbeek, Michal Drozdzal, and Adriana Romero Soriano. 2021. Instance-conditioned GAN. In *Advances in Neural Information Processing Systems*.

[15] Lucy Chai, Michael Gharbi, Eli Shechtman, Phillip Isola, and Richard Zhang. 2022. Any-resolution training for high-resolution image synthesis. In *Proceedings of the European Conference on Computer Vision*.

[16] Keshigeyan Chandrasegaran, Ngoc-Trung Tran, Alexander Binder, and Ngai-Man Cheung. 2022. Discovering Transferable Forensic Features for CNN-generated Images Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

[17] Keshigeyan Chandrasegaran, Ngoc-Trung Tran, and Ngai-Man Cheung. 2021. A closer look at Fourier spectrum discrepancies for CNN-generated images detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[18] K. Chandrasegaran, N. T. Tran, Y. Zhao, and N. M. Cheung. 2022. Revisiting label smoothing and knowledge distillation compatibility: What was missing?. In *Proceedings of the International Conference on Machine Learning*.

[19] Huiwen Chang, Han Zhang, Lu Jiang, Ce Liu, and William T Freeman. 2022. Maskgit: Masked generative image transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[20] Uttam Chauhan and Apurva Shah. 2021. Topic modeling using latent Dirichlet allocation: A survey. *ACM Computing Surveys (CSUR)* 54, 7 (2021), 1–35.
[21] Tianlong Chen, Yu Cheng, Zhe Gan, Jingjing Liu, and Zhangyang Wang. 2021. Data-efficient GAN training beyond (just) augmentations: A lottery ticket perspective. In *Advances in Neural Information Processing Systems*.
[22] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. 2020. StarGAN v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
[23] Min Jin Chong and David Forsyth. 2022. JoJoGAN: One shot face stylization. In *Proceedings of the European Conference on Computer Vision*. Springer.
[24] Laurie Clarke. 2022. When AI can make art – what does it mean for creativity? (2022). https://www.theguardian.com/technology/2022/nov/12/when-ai-can-make-art-what-does-it-mean-for-creativity-dall-e-midjourney
[25] Louis Clouâtre and Marc Demers. 2019. FIGR: Few-shot Image Generation with Reptile. *arXiv preprint arXiv:1901.02199* (2019).
[26] Yulai Cong, Miaoyun Zhao, Jianqiao Li, Sijia Wang, and Lawrence Carin. 2020. GAN memory with no forgetting. In *Advances in Neural Information Processing Systems*.
[27] Kaiwen Cui, Jiaxing Huang, Zhipeng Luo, Gongjie Zhang, Fangneng Zhan, and Shijian Lu. 2022. GenCo: Generative co-training for generative adversarial networks with limited data. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
[28] Kaiwen Cui, Yingchen Yu, Fangneng Zhan, Shengcai Liao, Shijian Lu, and Eric Xing. 2023. KD-DLGAN: Data limited image generation via knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[29] Mengyu Dai, Haibin Hang, and Xiaoyang Guo. 2022. Adaptive feature interpolation for low-shot image generation. In *Proceedings of the European Conference on Computer Vision*.
[30] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[31] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of North American Chapter of the Association for Computational Linguistics*.
[32] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*.
[33] Guanqi Ding, Xinzhe Han, Shuhui Wang, Xin Jin, Dandan Tu, and Qingming Huang. 2023. Stable attribute group editing for reliable few-shot image generation. *arXiv preprint arXiv:2302.00179* (2023).
[34] Guanqi Ding, Xinzhe Han, Shuhui Wang, Shuzhe Wu, Xin Jin, Dandan Tu, and Qingming Huang. 2022. Attribute group editing for reliable few-shot image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[35] Yuxuan Duan, Li Niu, Yan Hong, and Liqing Zhang. 2023. WeditGAN: Few-shot image generation via latent space relocation. *arXiv preprint arXiv:2305.06671* (2023).
[36] Ricard Durall, Margret Keuper, and Janis Keuper. 2020. Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[37] Joe Easton and Sarah Jacob. 2023. New Beatles song shows AI opportunity for music, Citigroup says. (2023). https://www.bloomberg.com/news/articles/2023-06-15/new-beatles-song-shows-ai-opportunity-for-music-citigroup-says?leadSource=uverify%20wall
[38] Patrick Esser, Robin Rombach, and Bjorn Ommer. 2021. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[39] Tiantian Fang, Ruoyu Sun, and Alex Schwing. 2022. DigGAN: Discriminator gradIent Gap regularization for GAN training with limited data. In *Advances in Neural Information Processing Systems*.
[40] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep network. In *Proceedings of the International Conference on Machine Learning*.
[41] Jonathan Frankle and Michael Carbin. 2019. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*.
[42] Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. 2023. DreamSim: Learning New Dimensions of Human Visual Similarity using Synthetic Data. *arXiv:2306.09344* (2023).
[43] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618* (2022).
[44] Rinon Gal, Or Patashnik, Haggai Maron, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. StyleGAN-NADA: CLIP-guided domain adaptation of image generators. *ACM Transactions on Graphics* 41, 4 (2022), 1–13.

[45] Giorgio Giannone, Didrik Nielsen, and Ole Winther. 2022. Few-shot diffusion models. *arXiv preprint arXiv:2205.15463* (2022).
[46] Giorgio Giannone and Ole Winther. 2022. SCHA-VAE: Hierarchical context aggregation for few-Shot generation. In *Proceedings of the International Conference on Machine Learning*.
[47] Rafael C Gonzales and Paul Wintz. 1987. *Digital image processing*. Addison-Wesley Longman Publishing Co., Inc.
[48] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems*.
[49] Yao Gou, Min Li, Yilong Lv, Yusen Zhang, Yuhang Xing, and Yujie He. 2023. Rethinking cross-domain semantic relation for few-shot image generation. *Applied Intelligence* (2023), 1–14.
[50] Timofey Grigoryev, Andrey Voynov, and Artem Babenko. 2022. When, why, and which pretrained GANs are useful?. In *International Conference on Learning Representations*.
[51] Zheng Gu, Wenbin Li, Jing Huo, Lei Wang, and Yang Gao. 2021. LoFGAN: Fusing local representations for few-shot image generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
[52] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. 2017. Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems*.
[53] Jiayi Guo, Chaofei Wang, You Wu, Eric Zhang, Kai Wang, Xingqian Xu, Humphrey Shi, Gao Huang, and Shiji Song. 2023. Zero-shot generative model adaptation via image-specific prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[54] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[55] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[56] Alex Hern. 2023. The AI tools that will write our emails, attend our meetings – and change our lives . (2023). https://www.theguardian.com/technology/2023/mar/21/the-ai-tools-that-will-write-our-emails-attend-our-meetings-and-change-our-lives
[57] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*.
[58] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. In *NeurIPS Deep Learning and Representation Learning Workshop*.
[59] Tobias Hinz, Matthew Fisher, Oliver Wang, and Stefan Wermter. 2021. Improved techniques for training single-image GANs. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*.
[60] Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. 2019. Flow++: Improving flow-based generative models with variational dequantization and architecture design. In *Proceedings of the International Conference on Machine Learning*.
[61] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*.
[62] Yan Hong, Li Niu, Jianfu Zhang, and Liqing Zhang. 2020. MatchingGAN: Matching-based few-shot image generation. In *IEEE International Conference on Multimedia and Expo*.
[63] Yan Hong, Li Niu, Jianfu Zhang, and Liqing Zhang. 2022. DeltaGAN: Towards diverse few-shot image generation with sample-specific delta. In *Proceedings of the European Conference on Computer Vision*.
[64] Yan Hong, Li Niu, Jianfu Zhang, and Liqing Zhang. 2022. Few-shot image generation using discrete content representation. In *Proceedings of the ACM International Conference on Multimedia*.
[65] Yan Hong, Li Niu, Jianfu Zhang, Weijie Zhao, Chen Fu, and Liqing Zhang. 2020. F2GAN: Fusing-and-filling GAN for few-shot image generation. In *Proceedings of the 28th ACM international conference on multimedia*. 2535–2543.
[66] George Hopkin. 2023. Microsoft's Azure OpenAI Service offers ChatGPT for business. *AI Magazine* (2023). https://aimagazine.com/articles/microsofts-azure-openai-service-offers-chatgpt-for-business
[67] Liang Hou. 2023. Regularizing label-augmented generative adversarial networks under limited data. *IEEE Access* 11 (2023), 28966–28976.
[68] Liang Hou, Huawei Shen, Qi Cao, and Xueqi Cheng. 2021. Self-supervised GANs with label augmentation. In *Advances in Neural Information Processing Systems*.
[69] Xingzhong Hou, Boxiao Liu, Fang Wan, and Haihang You. 2022. Exploiting knowledge distillation for few-shot image generation. https://openreview.net/forum?id=vsEi1UMa7TC
[70] Xingzhong Hou, Boxiao Liu, Shuai Zhang, Lulin Shi, Zite Jiang, and Haihang You. 2022. Dynamic weighted semantic correspondence for few-shot image generative adaptation. In *Proceedings of the ACM International Conference on Multimedia*.

[71] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[72] Jiaxing Huang, Kaiwen Cui, Dayan Guan, Aoran Xiao, Fangneng Zhan, Shijian Lu, Shengcai Liao, and Eric Xing. 2022. Masked generative adversarial networks are data-efficient generation learners. In *Advances in Neural Information Processing Systems*.
[73] Abdul Jabbar, Xi Li, and Bourahla Omar. 2021. A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys (CSUR)* 54, 8 (2021), 1–49.
[74] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. 2021. Deceive D: Adaptive pseudo augmentation for GAN training with limited data. In *Advances in Neural Information Processing Systems*.
[75] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. 2022. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*.
[76] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2020. Training generative adversarial networks with limited data. In *Advances in Neural Information Processing Systems*.
[77] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[78] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of StyleGAN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[79] Yuichi Kato, Masahiko Mikawa, and Makoto Fujisawa. 2023. Faster few-shot face image generation With features of specific group using pivotal tuning inversion and PCA. In *International Conference on Artificial Intelligence in Information and Communication*.
[80] Mahyar Khayatkhoei and Ahmed Elgammal. 2022. Spatial frequency bias in convolutional generative adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
[81] Seongtae Kim, Kyoungkook Kang, Geonung Kim, Seung-Hwan Baek, and Sunghyun Cho. 2022. DynaGAN: Dynamic few-shot adaptation of GANs to multiple domains. In *ACM Transactions on Graphics (SIGGRAPH Asia)*.
[82] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. 2021. Variational diffusion models. *Advances in Neural Information Processing Systems*.
[83] Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *International Conference on Learning Representations*.
[84] Diederik P Kingma, Max Welling, et al. 2019. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning* 12, 4 (2019), 307–392.
[85] Idan Kligvasser, Tamar Rott Shaham, Noa Alkobi, and Tomer Michaeli. 2022. BlendGAN: Learning and blending the internal distributions of single images by spatial image-identity conditioning. *arXiv preprint arXiv:2212.01589* (2022).
[86] Jing Yu Koh, Daniel Fried, and Ruslan Salakhutdinov. 2023. Generating images with multimodal language models. *arXiv preprint arXiv:2305.17216* (2023).
[87] Chaerin Kong, Jeesoo Kim, Donghoon Han, and Nojun Kwak. 2022. Few-shot image generation with mixup-based distance learning. In *Proceedings of the European Conference on Computer Vision*.
[88] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images.
[89] Vladimir Kulikov, Shahar Yadin, Matan Kleiner, and Tomer Michaeli. 2023. SinDDm: A single image denoising diffusion model. In *Proceedings of the International Conference on Machine Learning*.
[90] Rajiv Kumar and G Sivakumar. 2023. DEff-GAN: Diverse attribute transfer for few-shot image synthesis. *arXiv preprint arXiv:2302.14533* (2023).
[91] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. 2023. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[92] Nupur Kumari, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. 2022. Ensembling off-the-shelf models for GAN training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[93] Gihyun Kwon and Jong Chul Ye. 2023. One-shot adaptation of GAN in just one CLIP. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
[94] Tuomas Kynkäänniemi, Tero Karras, Miika Aittala, Timo Aila, and Jaakko Lehtinen. 2023. The role of ImageNet classes in Fréchet Inception Distance. In *International Conference on Learning Representations*.
[95] Hyuk-Gi Lee, Gi-Cheon Kang, Changhoon Jeong, Han-Wool Sul, and Byoung-Tak Zhang. 2021. C^3: Contrastive learning for cross-domain correspondence in few-shot image generation. *Controllable Generative Modeling in Language and Vision Workshop at NeurIPS*.
[96] Dongxu Li, Junnan Li, and Steven CH Hoi. 2023. Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing. *arXiv preprint arXiv:2305.14720* (2023).
[97] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597* (2023).

[98] Ke Li and Jitendra Malik. 2018. Implicit maximum likelihood estimation. *arXiv preprint arXiv:1809.09087* (2018).
[99] Lingxiao Li, Yi Zhang, and Shuhui Wang. 2022. The Euclidean space is evil: Hyperbolic attribute editing for few-shot image generation. *arXiv preprint arXiv:2211.12347* (2022).
[100] Tianqin Li, Zijie Li, Harold Rockwell, Amir Farimani, and Tai Sing Lee. 2022. Prototype memory and attention mechanisms for few-shot image generation. In *Proceedings of the Eleventh International Conference on Learning Representations*.
[101] Wenkuan Li, Wenyi Xu, Xubin Wu, Qianshan Wang, Qiang Lu, Tianxia Song, and Haifang Li. 2023. AMMGAN: Adaptive multi-scale modulation generative adversarial network for few-shot image generation. *Applied Intelligence* (2023), 1–19.
[102] Yanghao Li, Haoqi Fan, Ronghang Hu, Christoph Feichtenhofer, and Kaiming He. 2023. Scaling language-image pre-training via masking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[103] Yijun Li, Richard Zhang, Jingwan Lu, and Eli Shechtman. 2020. Few-shot image generation with elastic weight consolidation. In *Advances in Neural Information Processing Systems*.
[104] Ziqiang Li, Chaoyue Wang, Heliang Zheng, Jing Zhang, and Bin Li. 2022. FakeCLR: Exploring contrastive learning for solving latent discontinuity in data-efficient GANs. In *Proceedings of the European Conference on Computer Vision*.
[105] Ziqiang Li, Beihao Xia, Jing Zhang, Chaoyue Wang, and Bin Li. 2022. A comprehensive survey on data-efficient GANs in image generation. *arXiv preprint arXiv:2204.08329* (2022).
[106] Weixin Liang, Zixuan Liu, and Can Liu. 2020. Dawson: A domain adaptive few shot generation framework. *arXiv preprint arXiv:2001.00576* (2020).
[107] Bingchen Liu, Yizhe Zhu, Kunpeng Song, and Ahmed Elgammal. 2021. Towards faster and stabilized GAN training for high-fidelity few-shot image synthesis. In *International Conference on Learning Representations*.
[108] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. 2022. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[109] Haoming Lu, Hazarapet Tunanyan, Kai Wang, Shant Navasardyan, Zhangyang Wang, and Humphrey Shi. 2023. Specialist diffusion: Plug-and-play sample-efficient fine-tuning of text-to-image diffusion models to learn any unseen style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[110] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2018. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*.
[111] Yisroel Mirsky and Wenke Lee. 2021. The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)* 54, 1 (2021), 1–41.
[112] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. 2020. Freeze the discriminator: a simple baseline for fine-tuning GANs. *CVPR AI for Content Creation Workshop* (2020).
[113] Arnab Kumar Mondal, Piyush Tiwary, Parag Singla, and Prathosh AP. 2023. Few-shot cross-domain image generation via inference-time latent-code learning. In *The Eleventh International Conference on Learning Representations*.
[114] Jongbo Moon, Hyunjun Kim, and Jae-Pil Heo. 2023. Progressive few-shot adaptation of generative model with align-free spatial correlation. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
[115] Eric Nguyen, Michael Poli, Marjan Faizi, Armin Thomas, Callum Birch-Sykes, Michael Wornow, Aman Patel, Clayton Rabideau, Stefano Massaroli, Yoshua Bengio, et al. 2023. HyenaDNA: Long-range genomic sequence modeling at single nucleotide resolution. *arXiv preprint arXiv:2306.15794* (2023).
[116] Ngoc-Bao Nguyen, Keshigeyan Chandrasegaran, Milad Abdollahzadeh, and Ngai-Man Cheung. 2023. Re-thinking model inversion attacks against deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[117] Alex Nichol, Joshua Achiam, and John Schulman. 2018. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999* (2018).
[118] Alexander Quinn Nichol and Prafulla Dhariwal. 2021. Improved denoising diffusion probabilistic models. In *Proceedings of the International Conference on Machine Learning*.
[119] Yaniv Nikankin, Niv Haim, and Michal Irani. 2023. SinFusion: Training diffusion models on a single image or video. In *International Conference on Machine Learning*.
[120] Maria-Elena Nilsback and Andrew Zisserman. 2008. Automated flower classification over a large number of classes. In *Indian Conference on Computer Vision, Graphics & Image Processing*.
[121] Atsuhiro Noguchi and Tatsuya Harada. 2019. Image generation from small datasets via batch statistics adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
[122] Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. 2021. Few-shot image generation via cross-domain correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[123] Alan V Oppenheim, Alan S Willsky, Syed Hamid Nawab, and Jian-Jiun Ding. 1997. *Signals and systems*. Vol. 2. Prentice hall Upper Saddle River, NJ.

[124] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[125] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[126] Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. 2022. On Aliased Resizing and Surprising Subtleties in GAN Evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[127] Aniwat Phaphuangwittayakul, Yi Guo, and Fangli Ying. 2021. Fast adaptive meta-learning for few-shot image generation. *IEEE Transactions on Multimedia* 24 (2021), 2205–2217.

[128] Aniwat Phaphuangwittayakul, Fangli Ying, Yi Guo, Liting Zhou, and Nopasit Chakpitak. 2022. Few-shot image generation based on contrastive meta-learning generative adversarial network. *The Visual Computer* (2022), 1–14.

[129] Dinh Q Phung, Thi V Duong, Svetha Venkatesh, and Hung H Bui. 2005. Topic transition detection using hierarchical hidden Markov and semi-Markov models. In *Proceedings of the 13th annual ACM international conference on Multimedia*. 11–20.

[130] Piotr Porwik and Agnieszka Lisowska. 2004. The Haar-wavelet transform in digital image processing: its status and achievements. *Machine graphics and vision* 13, 1/2 (2004), 79–98.

[131] Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and Sundaraja S Iyengar. 2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–36.

[132] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*.

[133] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. 2019. On the spectral bias of neural networks. In *Proceedings of the International Conference on Machine Learning*.

[134] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with CLIP latents. *arXiv preprint arXiv:2204.06125* (2022).

[135] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *Proceedings of the International Conference on Machine Learning*.

[136] Reuters. 2023. Microsoft launches image-creation tool on Bing powered by OpenAI's tech. (2023). https://www.reuters.com/technology/microsoft-rolls-out-image-creator-bing-powered-by-openais-technology-2023-03-21/

[137] MIT Technology Review. 2023. Turbo-charging productivity in Asia: the economic benefits of generative AI. (2023). https://www.technologyreview.com/2023/07/05/1075792/turbo-charging-productivity-in-asia-the-economic-benefits-of-generative-ai/

[138] Douglas A Reynolds et al. 2009. Gaussian mixture models. *Encyclopedia of biometrics* 741, 659-663 (2009).

[139] Esther Robb, Wen-Sheng Chu, Abhishek Kumar, and Jia-Bin Huang. 2020. Few-shot adaptation of generative adversarial networks. *arXiv preprint arXiv:2010.11943* (2020).

[140] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[141] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*.

[142] Kevin Roose. 2023. Don't ban ChatgPt in schools. teach with it. (2023). https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html?searchResultPosition=60

[143] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[144] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. In *Advances in Neural Information Processing Systems*.

[145] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved techniques for training GANs. In *Advances in Neural Information Processing Systems*.

[146] Aditya Sanghi, Rao Fu, Vivian Liu, Karl DD Willis, Hooman Shayani, Amir H Khasahmadi, Srinath Sridhar, and Daniel Ritchie. 2023. CLIP-Sculptor: Zero-shot generation of high-fidelity and diverse shapes from natural language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[147] Claudio Filipi Gonçalves Dos Santos and João Paulo Papa. 2022. Avoiding overfitting: A survey on regularization methods for convolutional neural networks. *Comput. Surveys* 54, 10s (2022), 1–25.

[148] Kyle Sargent, Jing Yu Koh, Han Zhang, Huiwen Chang, Charles Herrmann, Pratul Srinivasan, Jiajun Wu, and Deqing Sun. 2023. Vq3d: Learning a 3d-aware generative model on imagenet. *arXiv preprint arXiv:2302.06833* (2023).
[149] Axel Sauer, Kashyap Chitta, Jens Müller, and Andreas Geiger. 2021. Projected GANs converge faster. In *Advances in Neural Information Processing Systems*.
[150] Divya Saxena and Jiannong Cao. 2021. Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)* 54, 3 (2021), 1–42.
[151] Divya Saxena, Jiannong Cao, Jiahao Xu, and Tarun Kulshrestha. 2023. Re-GAN: Data-efficient GANs training via architectural reconfiguration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[152] Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, and Aran Komatsuzaki. 2021. Laion-400m: Open dataset of clip-filtered 400 million image-text pairs.
[153] Katja Schwarz, Yiyi Liao, and Andreas Geiger. 2021. On the frequency bias of generative models. In *Advances in Neural Information Processing Systems*.
[154] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. 2019. SinGAN: Learning a generative model from a single natural image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
[155] Mohamad Shahbazi, Martin Danelljan, Danda Pani Paudel, and Luc Van Gool. 2022. Collapse by conditioning: Training class-conditional GANs with limited data. In *International Conference on Learning Representations*.
[156] Mohamad Shahbazi, Zhiwu Huang, Danda Pani Paudel, Ajad Chhatkuli, and Luc Van Gool. 2021. Efficient conditional GAN transfer with knowledge propagation across classes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[157] Jiachen Shi, Wenzhen Liu, Guoqiang Zhou, and Yuming Zhou. 2023. AutoInfo GAN: Toward a better image synthesis GAN framework for high-fidelity few-shot datasets via NAS and contrastive learning. *Knowledge-Based Systems* 276 (2023), 110757.
[158] Abhishek Sinha, Jiaming Song, Chenlin Meng, and Stefano Ermon. 2021. D2c: Diffusion-decoding models for few-shot conditional generation. In *Advances in Neural Information Processing Systems*.
[159] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*.
[160] Kihyuk Sohn, Huiwen Chang, José Lezama, Luisa Polania, Han Zhang, Yuan Hao, Irfan Essa, and Lu Jiang. 2023. Visual prompt tuning for generative transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[161] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv:2010.02502* (2020).
[162] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. 2018. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[163] Vadim Sushko, Jurgen Gall, and Anna Khoreva. 2021. One-shot GAN: Learning to generate samples from single images and videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[164] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[165] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*.
[166] Christopher TH Teo, Milad Abdollahzadeh, and Ngai-Man Cheung. 2023. Fair generative models via transfer learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
[167] Kowshik Thopalli, Rakshith Subramanyam, Pavan Turaga, and Jayaraman J Thiagarajan. 2023. Target-aware generative augmentations for single-shot adaptation. *arXiv preprint arXiv:2305.13284* (2023).
[168] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. 2021. On data augmentation for GAN training. *IEEE Transactions on Image Processing* 30 (2021), 1882–1897.
[169] Hung-Yu Tseng, Lu Jiang, Ce Liu, Ming-Hsuan Yang, and Weilong Yang. 2021. Regularizing generative adversarial networks under limited data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[170] Arash Vahdat and Jan Kautz. 2020. NVAE: A deep hierarchical variational autoencoder. In *Advances in Neural Information Processing Systems*.
[171] Aaron Van Den Oord, Oriol Vinyals, et al. 2017. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*.
[172] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 86 (2008), 2579–2605.
[173] Charles F Van Loan. 1976. Generalizing the singular value decomposition. *SIAM J. Numer. Anal.* 13, 1 (1976), 76–83.

[174] Sakshi Varshney, Vinay Kumar Verma, PK Srijith, Lawrence Carin, and Piyush Rai. 2021. Cam-GAN: Continual adaptation modules for generative adversarial networks. In *Advances in Neural Information Processing Systems*.

[175] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*.

[176] Chi Wang, Min Zhou, Tiezheng Ge, Yuning Jiang, Hujun Bao, and Weiwei Xu. 2023. CF-Font: Content fusion for few-shot font generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[177] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. 2019. Panet: Few-shot image semantic segmentation with prototype alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

[178] Weilun Wang, Jianmin Bao, Wengang Zhou, Dongdong Chen, Dong Chen, Lu Yuan, and Houqiang Li. 2022. SinDiffusion: Learning a diffusion model from a single natural image. *arXiv preprint arXiv:2211.12445* (2022).

[179] Xiaogang Wang and Xiaoou Tang. 2008. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 11 (2008), 1955–1967.

[180] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. 2020. MineGAN: Effective knowledge transfer from GANs to target domains with few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[181] Yaxing Wang, Abel Gonzalez-Garcia, Chenshen Wu, Luis Herranz, Fahad Shahbaz Khan, Shangling Jui, and Joost Van de Weijer. 2021. Minegan++: Mining generative models for efficient knowledge transfer to limited data domains. *arXiv preprint arXiv:2104.13742* (2021).

[182] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost Van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. 2018. Transferring GANs: Generating images from limited data. In *Proceedings of the European Conference on Computer Vision*.

[183] Yue Wang, Ran Yi, Ying Tai, Chengjie Wang, and Lizhuang Ma. 2022. CtlGAN: Few-shot artistic portraits generation with contrastive transfer learning. *arXiv preprint arXiv:2203.08612* (2022).

[184] Zhendong Wang, Yifan Jiang, Huangjie Zheng, Peihao Wang, Pengcheng He, Zhangyang Wang, Weizhu Chen, and Mingyuan Zhou. 2023. Patch diffusion: Faster and more data-efficient training of diffusion models. *arXiv preprint arXiv:2304.12526* (2023).

[185] Zhendong Wang, Huangjie Zheng, Pengcheng He, Weizhu Chen, and Mingyuan Zhou. 2023. Diffusion-GAN: Training GANs with diffusion. In *International Conference on Learning Representations*.

[186] Yuxin Wen, John Kirchenbauer, Jonas Geiping, and Tom Goldstein. 2023. Tree-Ring Watermarks: Fingerprints for Diffusion Images that are Invisible and Robust. *arXiv preprint arXiv:2305.20030* (2023).

[187] Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. 2023. Data collection and quality challenges in deep learning: A data-centric ai perspective. *The VLDB Journal* 32, 4 (2023), 791–813.

[188] Xintian Wu, Huanyu Wang, Yiming Wu, and Xi Li. 2023. D3T-GAN: Data-dependent domain transfer GANs for image generation with limited data. *ACM Transactions on Multimedia Computing, Communications and Applications* 19, 4 (2023), 1–20.

[189] Jiayu Xiao, Liang Li, Chaofei Wang, Zheng-Jun Zha, and Qingming Huang. 2022. Few shot generative model adaption via relaxed spatial structural alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[190] Tammy Xu. 2022. Google's new AI can hear a snippet of song—and then keep on playing. (2022). https://www.technologyreview.com/2022/10/07/1060897/ai-audio-generation/

[191] Ceyuan Yang, Yujun Shen, Yinghao Xu, Deli Zhao, Bo Dai, and Bolei Zhou. 2022. Improving GANs with a dynamic discriminator. In *Advances in Neural Information Processing Systems*.

[192] Ceyuan Yang, Yujun Shen, Yinghao Xu, and Bolei Zhou. 2021. Data-efficient instance generation from instance discrimination. In *Advances in Neural Information Processing Systems*.

[193] Ceyuan Yang, Yujun Shen, Zhiyi Zhang, Yinghao Xu, Jiapeng Zhu, Zhirong Wu, and Bolei Zhou. 2021. One-shot generative domain adaptation. *arXiv preprint arXiv:2111.09876* (2021).

[194] Mengping Yang, Saisai Niu, Zhe Wang, Dongdong Li, and Wenli Du. 2023. DFSGAN: Introducing editable and representative attributes for few-shot image generation. *Engineering Applications of Artificial Intelligence* 117 (2023), 105519.

[195] Mengping Yang, Zhe Wang, Ziqiu Chi, and Wenli Du. 2023. ProtoGAN: Towards high diversity and fidelity image synthesis under limited data. *Information Sciences* 632 (2023), 698–714.

[196] Mengping Yang, Zhe Wang, Ziqiu Chi, and Wenyi Feng. 2022. WaveGAN: Frequency-aware GAN for high-fidelity few-shot image generation. In *Proceedings of the European Conference on Computer Vision*.

[197] Mengping Yang, Zhe Wang, Ziqiu Chi, and Yanbing Zhang. 2022. FreGAN: Exploiting frequency components for training GANs under limited data. In *Advances in Neural Information Processing Systems*.

[198] Jordan Yaniv, Yael Newman, and Ariel Shamir. 2019. The face of art: landmark detection and geometric style in portraits. *ACM Transactions on Graphics* 38, 4 (2019), 1–15.

[199] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *Advances in Neural Information Processing Systems*.
[200] Chaojian Yu, Bo Han, Li Shen, Jun Yu, Chen Gong, Mingming Gong, and Tongliang Liu. 2022. Understanding robust overfitting of adversarial training and beyond. In *Proceedings of the International Conference on Machine Learning*.
[201] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. 2015. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365* (2015).
[202] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
[203] Junhai Zhai, Sufang Zhang, Junfen Chen, and Qiang He. 2018. Autoencoder and its various variants. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 415–419.
[204] Han Zhang, Zizhao Zhang, Augustus Odena, and Honglak Lee. 2020. Consistency regularization for generative adversarial networks. In *International Conference on Learning Representations*.
[205] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[206] Wentian Zhang, Haozhe Liu, Bing Li, Jinheng Xie, Yawen Huang, Yuexiang Li, Yefeng Zheng, and Bernard Ghanem. 2023. Dynamically masked discriminator for generative adversarial networks. *arXiv preprint arXiv:2306.07716* (2023).
[207] Yabo Zhang, Yuxiang Wei, Zhilong Ji, Jinfeng Bai, Wangmeng Zuo, et al. 2022. Towards diverse and faithful one-shot adaption of generative adversarial networks. In *Advances in Neural Information Processing Systems*.
[208] Zhixing Zhang, Ligong Han, Arnab Ghosh, Dimitris N Metaxas, and Jian Ren. 2023. Sine: Single image editing with text-to-image diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[209] Zicheng Zhang, Yinglu Liu, Congying Han, Tiande Guo, Ting Yao, and Tao Mei. 2022. Generalized one-shot domain adaptation of generative adversarial networks. In *Advances in Neural Information Processing Systems*.
[210] Miaoyun Zhao, Yulai Cong, and Lawrence Carin. 2020. On leveraging pretrained GANs for generation with limited data. In *Proceedings of the International Conference on Machine Learning*.
[211] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. 2020. Differentiable augmentation for data-efficient GAN training. In *Advances in Neural Information Processing Systems*.
[212] Yunqing Zhao, Keshigeyan Chandrasegaran, Milad Abdollahzadeh, and Ngai-Man Man Cheung. 2022. Few-shot image generation via adaptation-aware kernel modulation. In *Advances in Neural Information Processing Systems*.
[213] Yunqing Zhao, Henghui Ding, Houjing Huang, and Ngai-Man Cheung. 2022. A closer look at few-shot image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[214] Yunqing Zhao, Chao Du, Milad Abdollahzadeh, Tianyu Pang, Min Lin, Shuicheng Yan, and Ngai-Man Cheung. 2023. Exploring incompatible knowledge transfer in few-shot image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[215] Yunqing Zhao, Tianyu Pang, Chao Du, Xiao Yang, Ngai-Man Cheung, and Min Lin. 2023. A recipe for watermarking diffusion models. *arXiv preprint arXiv:2303.10137* (2023).
[216] Zhengli Zhao, Sameer Singh, Honglak Lee, Zizhao Zhang, Augustus Odena, and Han Zhang. 2021. Improved consistency regularization for GANs. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
[217] Zhengli Zhao, Zizhao Zhang, Ting Chen, Sameer Singh, and Han Zhang. 2020. Image augmentations for GAN training. *arXiv preprint arXiv:2006.02595* (2020).
[218] Chenxi Zheng, Bangzhen Liu, Huaidong Zhang, Xuemiao Xu, and Shengfeng He. 2023. Where is my spot? Few-shot image generation via latent subspace optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[219] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 6 (2017), 1452–1464.
[220] Jingyuan Zhu, Huimin Ma, Jiansheng Chen, and Jian Yuan. 2022. Few-shot image generation via masked discrimination. *arXiv preprint arXiv:2210.15194* (2022).
[221] Jingyuan Zhu, Huimin Ma, Jiansheng Chen, and Jian Yuan. 2022. Few-shot image generation with diffusion models. *arXiv preprint arXiv:2211.03264* (2022).
[222] Jingyuan Zhu, Huimin Ma, Jiansheng Chen, and Jian Yuan. 2023. Few-shot 3D shape generation. *arXiv preprint arXiv:2305.11664* (2023).
[223] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*.
[224] Peihao Zhu, Rameen Abdal, John Femiani, and Peter Wonka. 2022. Mind the gap: Domain gap control for single shot domain adaptation for generative adversarial networks. In *International Conference on Learning Representations*.