756 A PROOF OF THEOREM 1

Notation. Let \mathcal{X} denote the observation space and $\mathcal{Y} = \{1, -1\}$ the output space. Denote P_{XY} as the joint probability of the joint space of $\mathcal{X} \times \mathcal{Y}$ and assume a meta distribution μ and n domains $P_{XY}^{(1)}, \dots, P_{XY}^{(i)}, P_{XY}^{(n)}$ are i.i.d realizations from μ . A decision function is a function $f \in \mathcal{F} : \mathcal{X} \to \mathcal{Y}$ predicts $\hat{y}_i = f(x_i)$. We denote $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$ a loss function and define the generalization error of a decision function as

$$\mathcal{L}^{\mu}(f) = \mathbb{E}_{P_{XY} \sim \mu} \mathbb{E}_{(x,y) \sim P_{XY}}[l(f(x), y)]$$
(3)

Since we have no access to μ and all the realizations $P_{XY}^{(1)}, \dots, P_{XY}^{(i)}, P_{XY}^{(n)}$ but sampled images from these realizations, we can derive an empirical error:

$$\hat{\mathcal{L}}^{\mu}(f) = \sum_{i=1}^{n} \sum_{j=1}^{m} l(f(x_{ij}), y_{ij})$$
(4)

1779 It's easy to see that when $n \to \infty, m \to \infty, \hat{\mathcal{L}}^{\mu}(f)$ converges to $\mathcal{L}^{\mu}(f)$, which gives the intuitive sense that increasing m and n gives us better-approximated solutions.

To prove Theorem [], we use a modified version of the standard empirical Rademacher complexity bound that weakens the i.i.d assumption to an independence assumption Mohri et al. (2018).

Theorem 2 For distribution $P^{(1)}, \dots, P^{(n)}$ independent sampled from meta-distribution μ , and *1*-Lipschitz loss $l(\cdot, \cdot)$ taking values in [0, 1], the following holds with confidence at least $1 - \delta$,

$$\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}_{P^{(j)}}(f) \le \frac{1}{n}\sum_{j=1}^{n}\hat{\mathcal{L}}_{P^{(j)}}(f) + 2\mathcal{R}_{mn}(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2mn}}$$
(5)

where $\hat{\mathcal{L}}_{P^{(j)}}(f)$ is losses on empirical set $S_{P^{(j)}}$ i.i.d. drawn from $P^{(j)}$.

Proof 1 Let $S = \bigcup_{i=1}^{n} S_{P^{(i)}}$ and

$$\Phi(S) = \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} (\mathcal{L}_{P^{(j)}}(f) - \hat{\mathcal{L}}_{P^{(j)}}(f))$$
(6)

which satisfies the bounded differences property required by McDiarmid's inequality, which implies that with confidence at least $1 - \frac{1}{2}\delta$ that

 $\Phi(S) \le \mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \left[\Phi(S) \right] + \sqrt{\frac{\ln(2/\delta)}{2mn}}$ (7)

Then we can bound the expected value of $\Phi(S)$

$$\mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \left[\Phi(S) \right] \tag{8}$$

$$= \mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} (\mathcal{L}_{P^{(j)}}(f) - \hat{\mathcal{L}}_{P^{(j)}}(f)) \right]$$
(9)

$$= \mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} \left(\mathbb{E}_{S'_{P^{(j)}} \sim P^{(j)}} \left[\frac{1}{m} \sum_{i=1}^{m} l(f(x'_{ij}), y'_{ij}] - \frac{1}{m} \sum_{i=1}^{m} l(f(x_{ij}), y_{ij}) \right) \right]$$
(10)

$$\leq \mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \mathbb{E}_{S'_{P^{(1:n)}} \sim P^{(1:n)}} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{m} \sum_{i=1}^{m} l(f(x'_{ij}), y'_{ij}) - l(f(x_{ij}), y_{ij}) \right]$$
(11)

$$= \mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \mathbb{E}_{S'_{P^{(1:n)}} \sim P^{(1:n)}} \mathbb{E}_{\sigma} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{m} \sum_{i=1}^{m} \sigma_{ij}(f(x'_{ij}), y'_{ij}) - l(f(x_{ij}), y_{ij}) \right]$$
(12)

$$\leq \mathbb{E}_{S'_{P^{(1:n)}} \sim P^{(1:n)}} \mathbb{E}_{\sigma} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{m} \sum_{i=1}^{m} \sigma_{ij}(f(x'_{ij}), y'_{ij}) \right]$$

$$(13)$$

$$+ \mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \mathbb{E}_{\sigma} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{m} \sum_{i=1}^{m} -\sigma_{ij}(f(x_{ij}), y_{ij}) \right]$$
(14)

$$= 2\mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \mathbb{E}_{\sigma} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{m} \sum_{i=1}^{m} \sigma_{ij} l(f(x_{ij}), y_{ij}) \right]$$
(15)

$$= 2\mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \left[\mathcal{R}_{mn}(\mathcal{F}) \right]$$
(16)

Following McDiarmid's inequality, we know that with confidence at least $1 - \frac{1}{2}\delta$,

$$2\mathbb{E}_{S_{P^{(1:n)}}\sim P^{(1:n)}}\left[\mathcal{R}_{mn}(\mathcal{F})\right] \le 2\mathcal{R}_{mn}(\mathcal{F}) + 2\sqrt{\frac{\ln(2/\delta)}{2mn}}$$
(17)

Finally, we have

$$\Phi(S) = \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{j=1}^{n} (\mathcal{L}_{P^{(j)}}(f) - \hat{\mathcal{L}}_{P^{(j)}}(f))$$
(18)

$$\leq \mathbb{E}_{S_{P^{(1:n)}} \sim P^{(1:n)}} \left[\Phi(S) \right] + \sqrt{\frac{\ln(2/\delta)}{2mn}} \tag{19}$$

$$\leq 2\mathcal{R}_{mn}(\mathcal{F}) + 2\sqrt{\frac{\ln(2/\delta)}{2mn}} + \sqrt{\frac{\ln(2/\delta)}{2mn}}$$
(20)

$$=2\mathcal{R}_{mn}(\mathcal{F})+3\sqrt{\frac{\ln(2/\delta)}{2mn}}$$
(21)

Thus,

$$\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}_{P^{(j)}}(f) \le \frac{1}{n}\sum_{j=1}^{n}\hat{\mathcal{L}}_{P^{(j)}}(f) + 2\mathcal{R}_{mn}(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2mn}}$$
(22)

which completes the proof.

Then we can derive the generalization bound with standard empirical Rademacher complexity bound Li et al. (2022a).

Theorem 3 For a 1-Lipschitz loss l, with confidence at least $1 - 2\delta$ and for all $f \in \mathcal{F}$, we have

$$\mathcal{L}^{\mu}(f) \leq \hat{\mathcal{L}}^{\mu}(f) + 2\mathcal{R}_{mn}(\mathcal{F}) + 2\mathcal{R}_{n}(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2mn}} + 3\sqrt{\frac{\ln(2/\delta)}{n}}$$

where $\mathcal{R}(\mathcal{F})$ standard empirical Rademacher complexity on function class \mathcal{F} .

Now we show that both the number of domains n and the number of images observed from each domain m is negatively correlated to the upper bound of generalization error.

Proof 2 Let $P = \{p^{(1)}, \dots, p^{(n)}\}$ be a set of *n* domain distribution i.i.d. sampled from ϵ . Define

$$\Phi(P) = \sup_{f \in \mathcal{F}} \mathcal{L}^{\epsilon}(f) - \frac{1}{n} \sum_{j=1}^{n} \mathcal{L}_{p^{(j)}}(f)$$
(23)

We construct P' by replacing any $p^{(j)} \in P$ with $p' \sim \mu$, then we have $|\Phi(P) - \Phi(P')| \leq 1/n$. Thus, McDiarmid's inequality tells us that with confidence at least $1 - \frac{1}{2}\delta$

$$\Phi(P) \le \mathbb{E}_{P^{(1:n)} \sim \mu} \left[\Phi(P) \right] + \sqrt{\frac{\ln(2/\delta)}{2n}}$$
(24)

Following the proof techniques in Theorem 2 we bound the expected value of $\Phi(P)$

$$\mathbb{E}_{P^{(1:n)} \sim \mu} \left[\Phi(P) \right] \tag{25}$$

$$= \mathbb{E}_{P^{(1:n)} \sim \mu} \left[\sup_{f \in \mathcal{F}} \left(\mathbb{E}_{q \sim \mu} [\mathcal{L}_q(f)] - \frac{1}{n} \sum_{j=1}^n \mathcal{L}_{p^{(j)}}(f) \right) \right]$$
(26)

$$\leq 2\mathbb{E}_{P^{(1:n)}\sim\mu}\mathbb{E}_{(x_j,y_j)\sim p^{(j)}}\left[\mathcal{R}_n(\mathcal{F})\right]$$
(27)

McDiarmid's inequality can be used to say with confidence $1 - \frac{1}{2}\delta$ that

$$2\mathbb{E}_{P^{(1:n)}\sim\mu}\mathbb{E}_{(x_j,y_j)\sim p^{(j)}}\left[\mathcal{R}_n(\mathcal{F})\right] \le 2\mathcal{R}_n(\mathcal{F}) + 2\sqrt{\frac{\ln(2/\delta)}{2n}}$$
(28)

Thus, we have

$$\Phi(P) = \sup_{f \in \mathcal{F}} \mathcal{L}^{\epsilon}(f) - \frac{1}{n} \sum_{j=1}^{n} \mathcal{L}_{p^{(j)}}(f)$$
(29)

$$\leq \mathbb{E}_{P^{(1:n)} \sim \mu} \left[\Phi(P) \right] + \sqrt{\frac{\ln(2/\delta)}{2n}} \tag{30}$$

$$\leq 2\mathcal{R}_n(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2n}} \tag{31}$$

With Theorem 2 we have with confidence at least $1 - \delta$ that,

$$\frac{1}{n}\sum_{j=1}^{n}\mathcal{L}_{p^{(j)}}(f) \le \hat{\mathcal{L}}^{\mu}(f) + 2\mathcal{R}_{nm}(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2nm}}$$
(32)

Finally, we have

$$\sup_{f \in \mathcal{F}} \mathcal{L}^{\epsilon}(f) \le \frac{1}{n} \sum_{j=1}^{n} \mathcal{L}_{p^{(j)}}(f) + 2\mathcal{R}_{n}(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2n}}$$
(33)

 $\leq 2\mathcal{R}_{nm}(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2nm}} + 2\mathcal{R}_n(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2n}}$ (34)

which completes the proof.

Then we prove our Theorem 1



Figure 6: Examples of synthetic images conditioned on novel domain knowledge from LLM. The first two columns (i.e. Caltech101 and VOC2017) are selected from VLCS datasets while the rest three columns are images generated based on the novel domains (i.e. fairytale, etc) provided by LLMs

Proof 3 With confidence at least
$$1 - 2\delta$$
 and for all $f \in \mathcal{F}$, we have

$$\mathcal{L}^{\mu}(f) - \hat{\mathcal{L}}^{\mu'}(f) = \mathcal{L}^{\mu}(f) - \mathcal{L}^{\mu'}(f) + \mathcal{L}^{\mu'}(f) - \hat{\mathcal{L}}^{\mu'}(f)$$
(35)

With Theorem 3, we have

$$\mathcal{L}^{\mu}(f) - \mathcal{L}^{\mu'}(f) + \mathcal{L}^{\mu'}(f) - \hat{\mathcal{L}}^{\mu'}(f)$$
(36)

$$\leq 2\mathcal{R}_{mn}(\mathcal{F}) + 2\mathcal{R}_n(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2mn}} + 3\sqrt{\frac{\ln(2/\delta)}{n}} + \mathcal{L}^{\mu}(f) - \mathcal{L}^{\mu'}$$
(37)

$$\leq 2\mathcal{R}_{mn}(\mathcal{F}) + 2\mathcal{R}_n(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2mn}} + 3\sqrt{\frac{\ln(2/\delta)}{n}} + \sup_f |\mathcal{L}^{\mu}(f) - \mathcal{L}^{\mu'}|$$
(38)

With the assumption that $D(\mu, \mu') = \sup_f |\mathcal{L}^{\mu}(f) - \mathcal{L}^{\mu'}| \le \epsilon$, we have

$$\mathcal{L}^{\mu}(f) - \hat{\mathcal{L}}^{\mu'}(f) \tag{39}$$

$$\leq 2\mathcal{R}_{mn}(\mathcal{F}) + 2\mathcal{R}_n(\mathcal{F}) + 3\sqrt{\frac{\ln(2/\delta)}{2mn}} + 3\sqrt{\frac{\ln(2/\delta)}{n}} + \epsilon$$
(40)

which finishes the proof.

B VISUALIZATION

We provide more examples of synthetic images conditioned on novel domain knowledge from LLM. We present in Figure 6 the synthetic images of VLCS datasets.

C PITFALL OF TEXT-TO-IMAGE GENERATION MODELS

971 Text-to-image generation models are by nature noisy as no strict control can be achieved. We present some pitfalls (commonly reported by the community) that will insert noise and influence the training

			1 3-0-8			
973		A DECEMBER OF				
974						
975						
976	ambiguous	ACTING CALL AND THE CALL AND A CALL				
977	classification	A DESCRIPTION OF THE OWNER OF THE				
978	classification					
979		class: horse	class: horse	class: house	class: guitar	
980		domain: castle	domain: castle	domain: street	domain: beach	
981		grounds	grounds	markets		
982						
983						
984			A COMMENT			
985	missing			12332		
986	object					
987	of interest		16 the second			
988						
989		class: car	class: car	class: car	class: dog	
990		domain: airport	domain: luxury	domain: night	domain: art deco	
991			estate	club		
992					A COMPANY AND A COMPANY	
002				0/	the state	
999 997						
995	distorted			OF		
996	torso		Θ			
997	10130					
998		class: giraffe	class: scissors	class: scissors	Class: horse	
999		domain: ancient	domain:	domain: chalk art	domain: castle	
1000			minimalism		grounds	
1004						
1001						
1001	distorted or					
1002 1003 1004	distorted or more/less					
1002 1003 1004 1005	distorted or more/less limps or					
1002 1003 1004 1005 1006	distorted or more/less limps or fingers					
1002 1003 1004 1005 1006 1007	distorted or more/less limps or fingers					
1002 1003 1004 1005 1006 1007 1008	distorted or more/less limps or fingers	class: horse	class: elephant	class: person	class: person	
1002 1003 1004 1005 1006 1007 1008 1009	distorted or more/less limps or fingers	class: horse domain: arctic	class: elephant domain: cubistic	class: person domain: cityview	class: person domain: office	
1002 1003 1004 1005 1006 1007 1008 1009 1010	distorted or more/less limps or fingers	class: horse domain: arctic	class: elephant domain: cubistic	Class: person domain: cityview	class: person domain: office	
1007 1003 1004 1005 1006 1007 1008 1009 1010 1011	distorted or more/less limps or fingers	class: horse domain: arctic Figure 7: Exa	class: elephant domain: cubistic	class: person domain: cityview f synthetic images.	class: person domain: office	
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012	distorted or more/less limps or fingers	class: horse domain: arctic Figure 7: Exa model. We show in	class: elephant domain: cubistic amples of pitfalls of Figure 7 where eac	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure 7 where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure [] where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	:h
1007 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure 7 where each domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch
1007 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure [] where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	:h
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure [] where each domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch.
1007 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure 7 where each domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch.
1007 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure 7 where each domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch:
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure ? where each domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure [] where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch.
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure [7] where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	class: person domain: office	ch.
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure [] where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	Elass: person domain: office	ch
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure [] where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	Eass: person domain: office	ch.
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025	distorted or more/less limps or fingers of a generalizable image is the corres	class: horse domain: arctic Figure 7: Exa model. We show in sponding class and c	class: elephant domain: cubistic amples of pitfalls of Figure ? where eac domain.	class: person domain: cityview f synthetic images. ch row is a type of p	Elass: person domain: office	ch